A TALE OF TWO ENZYMES: IDENTIFICATION OF AN UNKNOWN LIGAND BOUND TO CYTOCHROME P450 2A13 AND UNDERSTANDING SUBSTRATE SELECTIVITY OF CYTOCHROME P450 2E1

BY

Patrick R. Porubsky

B.S., Washburn University, 2006

Submitted to the Department of Medicinal Chemistry and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirement for the degree of Master's of Science.

Dissertation Committee:

_____
Major Professor

Committee

_____

_____

Date Defended:_____

Page left intentionally blank

The Thesis Committee of Patrick R. Porubsky certifies that this is the approved version of the following thesis:

A TALE OF TWO ENZYMES: IDENTIFICATION OF AN UNKNOWN LIGAND BOUND TO CYTOCHROME P450 2A13 AND UNDERSTANDING SUBSTRATE SELECTIVITY OF CYTOCHROME P450 2E1

Dissertation Committee:

_____
Major Professor

Committee

_____

_____

Date Defended:_____

Page left intentionally blank

Abstract

Cytochrome P450 (CYP) is the predominate superfamily of enzymes responsible for Phase I metabolism of drugs and other xenobiotics. Understanding the structural reasons for the substrate selectivity of these enzymes is important for both pharmacological and toxicological reasons. Two isoforms of interest from this enzyme superfamily that are CYP2A13 and CYP2E1.

Cytochrome P450 2A13 (CYP2A13) is a lung specific enzyme known to activate the potent tobacco procarcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) into two carcinogenic metabolites. CYP2A13 has been crystallized and X-ray diffraction experiments illuminated the structure of this enzyme, but with an unknown ligand present in the enzyme active site. This unknown ligand was suspected to be indole but a selective method had to be developed to differentiate among indole and its metabolites in the protein sample. We successfully modified a microbiological colorimetric assay to spectrophotometrically differentiate between indole and a number of possible indole metabolites in nanomolar concentrations by derivatization with *p*-dimethylaminocinnamaldehyde (DMACA). Further differentiation of indoles was made by mass spectrometry (HPLC–UV/vis–MS/MS) utilizing the chromophore generated in the DMACA conjugation as a UV signature for HPLC detection. The ligand in the crystallized protein was unambiguously identified as unsubstituted indole, which facilitated refinement of two alternate conformations of indole in the CYP2A13 crystal structure active site.

Human cytochrome P450 2E1 (CYP2E1) is a xenobiotic metabolizing enzyme that is highly conserved among mammals. In addition to small molecular weight exogenous drugs like the analgesic acetaminophen and the volatile anesthetic halothane, CYP2E1 is also involved in endogenous fatty acid metabolism. To more fully understand the structural factors that contribute to the substrate selectivity of CYP2E1, it has been cocrystallized with two structurally different heme-binding compounds: indazole, a small molecular weight inhibitor and ω-imidazolyl-decanoic acid, a fatty acid analog. Comparison of the CYP2E1 structures shows that only small side chain movements are required for the accommodation of the much larger fatty acid analog. Rotation of the side chain of F298 causes a change in the active site volume from 190 $Å^3$ in the indazole-bound structure to 440 $Å^3$ in the ω-imidazolyl-decanoic acid-bound structure. Future work will be focused on cocrystal structures of CYP2E1 with both longer and shorter chain analogs to better understand the ability of the enzyme to metabolize a variety of fatty acids substrates.

Acknowledgements

The author would like to thank many individuals and groups that have supported the work described in this thesis. Specific thanks to Brian Smith for the purified CYP2A13 used for the indole assay work. Dr. Kathleen Meneely deserves thanks for helpful collaboration in the pursuit of the first crystal structures of CYP2E1 (each of us investigating a different ligand bound structure). General thanks to the Scott lab members past and present including Brian Smith, Dr. Kathy Meneely, Dr. Lena Zaitzeva, Natasha DeVore, Linda Blake, Dr. Agnes Walsh, Melanie Blevins, and Eric Carrillo. The *pGEM 4-h2E1* plasmid was a gift from Dr. M. Ingelman-Sundberg (Karolinska Institute). Dr. Scott Lovell (The University of Kansas) deserves acknowledgment for crystallization advice and for coordinating collaboration with Dr. Kevin Battaile (Advanced Photon Source) who collected and processed the CYP2E1/ω-imidazolyl-decanoic acid crystal data. Thanks to Dr. Todd Williams (The University of Kansas) for guidance in the design of LC-MS/MS experiments. Thanks to the lab of Dr. Jane Aldrich (The University of Kansas) for the use of hardware essential for the organic chemistry discussed.

Funding of this work was provided by the NIH grants KINBRE (RR016475), NIGMS RO1 (GM076343), and COBRE (RR017708). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental

Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

I would like to thank my committee of Dr. Todd Williams, Dr. Audrey Lamb, and my research advisor Dr. Emily Scott. Additional thanks must be given to Dr. Scott for her excellent guidance over the past years.

Last and most importantly I would like to thank my wife, Michelle, for her support and patience throughout my graduate studies.

Table of Contents Page

List of Figures

List of Tables

3.2 List of Schemes

Page left intentionally blank

Chapter 1.

Introduction

Cytochrome P450: The Superfamily

The metabolism of drugs and other xenobiotics typically involves two classes of enzymes. These classes are referred to as Phase I and Phase II drug metabolizing enzymes. The Phase I enzymes are involved in either unmasking or generating functional groups to increase the water solubility of a compound. The Phase II enzymes are responsible for conjugating the substance to endogenous compounds such as sulfate, glutathione, and glucuronic acid. The most prevalent class of enzymes involved in Phase I metabolism is cytochrome P450 (CYP).

The CYP superfamily of enzymes was originally observed by Garfinkel (1) and Klingenberg (2) and subsequently named for its characteristic absorbance at 450 nm (Figure 1.1) in the presence of reducing agent (sodium dithionite) and carbon carbon monoxide (3-6). This spectral feature distinguishes CYP from other heme containing proteins that ligate heme with histidine. Originally, this superfamily was believed to be a single enzyme, but through induction and enzyme isolation it was shown that the entity with the characteristic absorbance peak was actually several related CYP isozymes (7-9). The CYP superfamily of enzymes has monooxygenase activity and has been shown to catalyze many different reactions, including the dealkylation of hetero-atoms (N, O, and S), aliphatic and aromatic hydroxylation, and olefin epoxidation (10). All of these reactions occur between the endogenous or

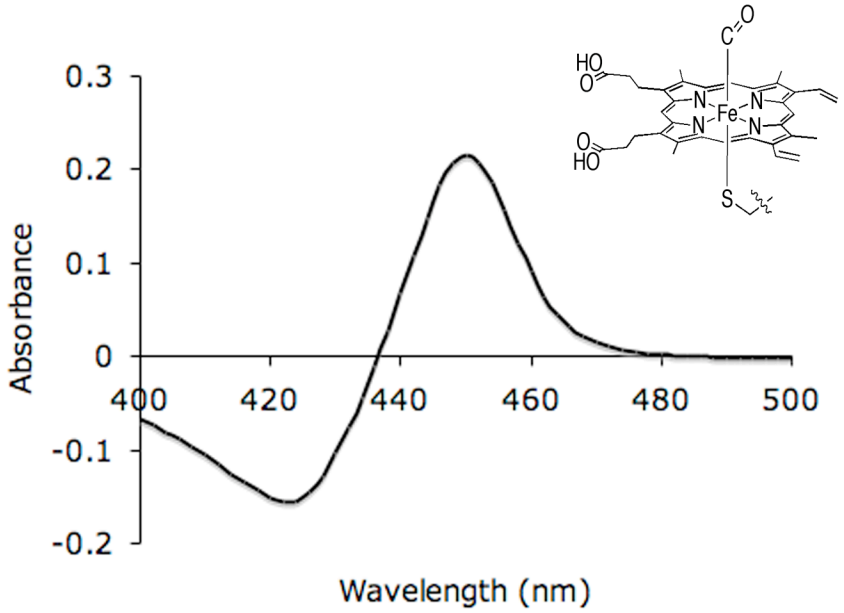


Figure 1.1: Characteristic carbon monoxide difference spectrum. Inset is the structure of the iron protoporphyrin IX bound with carbon monoxide and the thiolate donated by the CYP polypeptide.

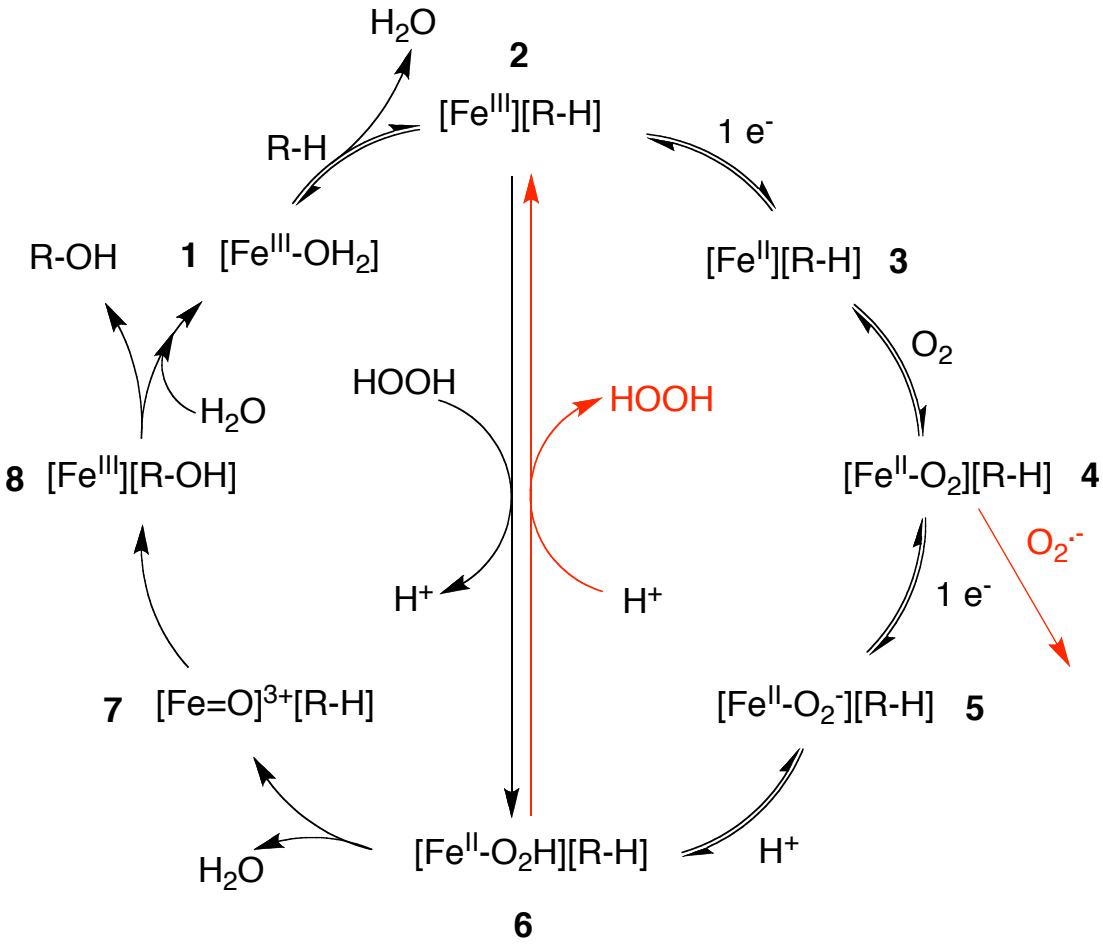
xenobiotic substrate and an activated oxygen/iron heme thiolate complex generated in the CYP active site.

The CYP Catalytic Process

The catalytic process of CYP is driven by reducing equivalents donated by NADH or NADPH. The electrons from the reducing equivalents are shuttled to CYP by a redox partner. The identity of the redox partner varies among CYPs. Cytosolic bacterial CYPs, as well as mitochondrial CYPs in eukaryotes, use the redox partners of putidaredoxin and adrenodoxin. The microsomal CYPs in eukaryotes, including those responsible for drug metabolism, utilize NADPH-dependent P450 oxidoreductase as a redox partner. This electron transport facilitates the reductive cleavage of molecular oxygen that is responsible the CYP oxidative power.

The CYP catalytic cycle (Scheme 1.1) starts with a resting state with ferric heme iron complexed with water (11). Substrate binding typically displaces the water molecule from the sixth coordinate position of iron. The conversion to five coordinate iron changes the iron oxidative potential facilitating the first electron transport (12) (Scheme 1.1 **3**). The newly formed ferrous iron/substrate bound form of the enzyme binds molecular oxygen in a coordinate covalent manner (Scheme 1.1 **4**). An additional one-electron reduction occurs generating a ferrous peroxyoxo-anion complex (Scheme 1.1 **5**). Protonation yields ferrous hydroperoxy species (Scheme 1.1 **6**), which upon loss of water generates the proposed highly reactive iron oxo species (Scheme 1.1 **7**). The exact identity of this iron oxo species is still under investigation but it is proposed that this species abstracts a hydrogen atom from the

Scheme 1.1: CYP catalytic cycle. Steps shown in red represent uncoupled pathways.



substrate. This generates a radical substrate and an iron hydroxy radical. The two radicals react to generate the oxidized product and convert the protein back to its initial ferric state. Several of the steps of the catalytic cycle have been confirmed crystallographically (11,13-15).

In some cases, half of the catalytic cycle can be bypassed by the addition of hydrogen peroxide (16) or some other peroxide source (for example cumin hydroperoxide (17)). This pathway is known as the hydrogen peroxide shunt (black arrow in center of Scheme 1.1) and is a path directly from ferric iron to the ferrous hydroperoxide species. The reverse of the hydrogen peroxide shunt is one of the two ways that this catalytic process can be "uncoupled" (red arrow in the center of Scheme 1.1). This "uncoupling" is in reference to when reducing equivalents are consumed but substrate is not oxidized because of an additional route out of the catalytic cycle. The other path of uncoupling is the decomposition of complex **4** of Scheme 1.1 to release a superoxide radical (red arrow on right in Scheme 1.1). These products contribute to cellular oxidative stress through lipid peroxidation, protein oxidation, and DNA damage.

CYP Nomenclature: Families and Subfamilies

The CYP superfamily is organized into families and subfamilies based on amino acid sequence identity. The recommended nomenclature used begins with CYP (**Cy**tochrome **P**450) followed an Arabic number and a letter (18). The Arabic number and letter represent the specific family and subfamily respectively. In most cases the CYP protein families are >40% percent identical, but there are exceptions to

this rule. The subfamilies (in the mammalian system) are generally >55% identical, but as low as 46% when subfamilies contain proteins from distant species (18). The P450 Nomenclature Committee officially designates these family and subfamily names. Finally, a terminal Arabic number designates an individual isozyme.

The families that have the greatest impact on human xenobiotic metabolism are CYP1, CYP2, and CYP3. Figure 1.2 illustrates the percent contribution of CYPs involved in the metabolism of clinically used drugs. Three isozymes, namely 3A4, 2D6, and 2C9, account for the metabolism of ~75% of the clinically used drugs (19). Enzymes known for the activating carcinogens include CYP1A1, CYP1A2, CYP2A6, CYP2E1, and CYP3A4 (20). Since 2000, investigating the structure of these mammalian membrane enzymes by X-ray crystallography has become an active field of research, but the basic CYP fold was originally determined from soluble bacterial isoforms.

CYP Structures

The first crystal structure from the CYP superfamily of enzymes was CYP101 ($P450_{cam}$) from the soil bacteria *Pseudomonas putida* (15). This enzyme is known for its metabolism of camphor hence the common name $P450_{cam}$. The general secondary structure features are primarily α-helical (named alphabetically) and a small amount of β-sheet (named $\beta_{\#-\#}$ with the first # indicating the sheet system and the second indicating which strand in the particular sheet system) (Figure 1.3). This naming scheme was preserved in future structures with additional helices given the

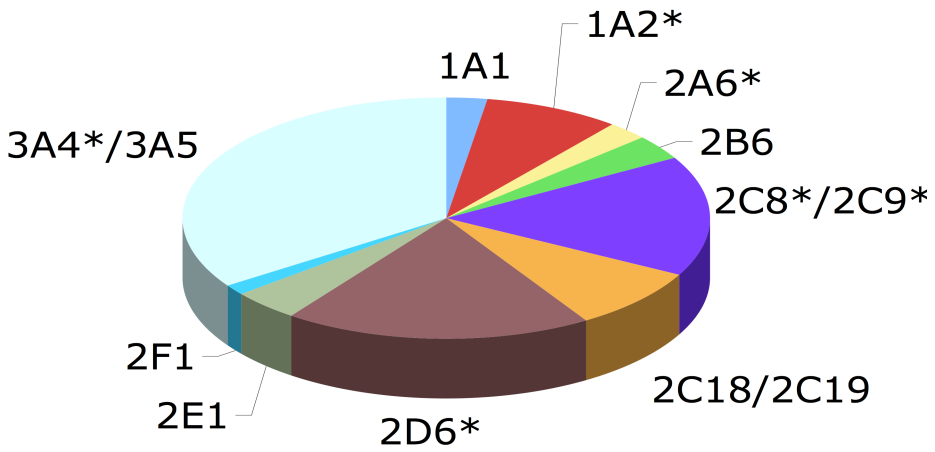


Figure 1.2: CYP isoforms metabolizing clinically used drugs (19). The size of the slices on the pie chart indicates the percentage contribution of the isozymes to the CYP catalyzed drug metabolism. Asterisks indicate isoforms whose crystal structures were solved prior to this work.

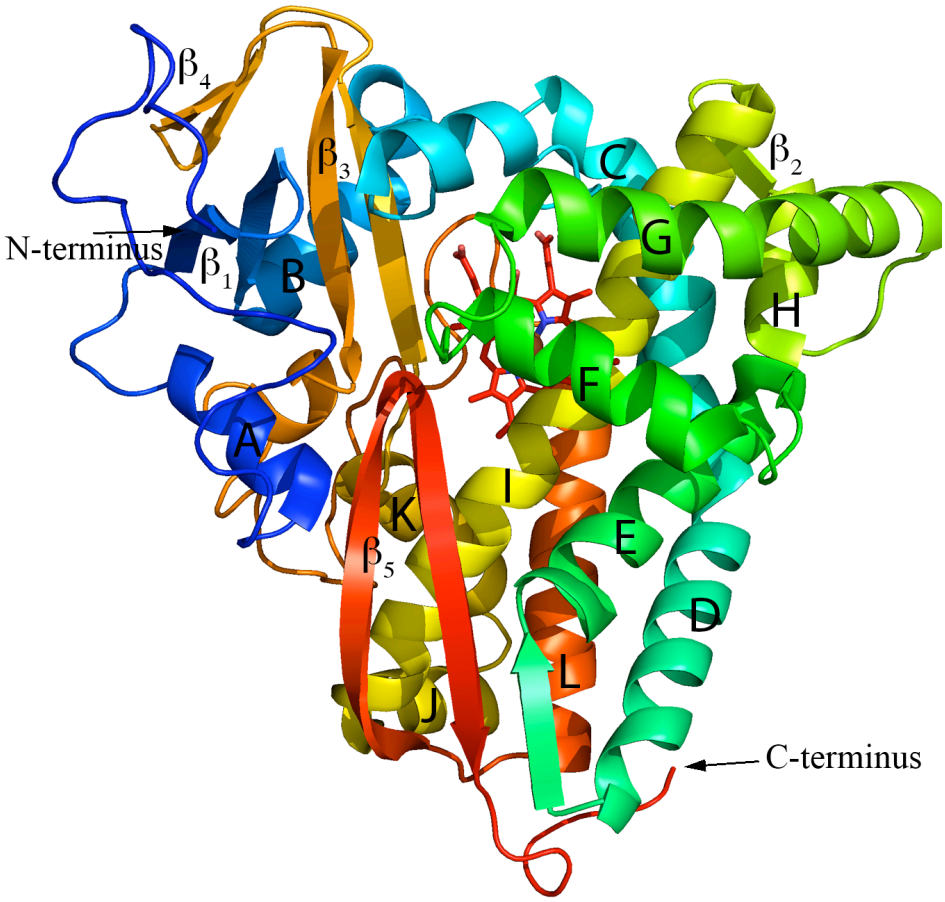


Figure 1.3: Overall structure of CYP101 (PDB code 2CPP (15)) colored from N-terminus (blue) to C-terminus (red).

prime nomenclature. For example, the helices F′ and G′ are helices in mammalian CYPs between helices F and G and that were not present in the original CYP101 structure. The CYP101 structure also confirmed the iron heme thiolate complex in the CYP active site.

The first mammalian CYP whose structure was solved was a mutant form of CYP2C5 (a rabbit CYP). This enzyme had N-terminal transmembrane residues 3-20 deleted to increase solubility and decrease aggregation, a 4xHis tag added to the C-terminus to assist in purification, and internal mutations (a chimeric form with CYP2C3 called CYP2C5/3LVdH) (21). This crystal structure showed a protein with an overall similar fold to the cytosolic bacterial forms previously crystallized. Compared with CYP101, the CYP2C5 structure has less β-sheet secondary structure, but had a new hydrophobic surface at the N-terminus of the protein thought to interact with the membrane. This structure was ground breaking, showing that a mammalian membrane CYP can be crystallized.

Shortly thereafter studies performed on another rabbit P450, CYP2B4, showed the dramatic conformational changes allowing substrate binding. The first structure (22) of CYP2B4 was in a homo-dimeric form with a histidine of one molecule binding to the heme iron of another and *vice-versa*. This interaction yielded protein in an open conformation with a cleft leading directly to the active site from the protein surface and allowing the dimerization interactions. When the coordinating histidine was mutated and an inhibitor bound in the active site, the protein crystallized in a closed conformation (23). This series of studies demonstrated the extensive

flexibility of the CYP structure illustrating an induced fit type of protein/ligand interaction.

Since the crystallization of these two non-human mammalian CYPs, extensive effort has been placed on solving the crystal structures of human drug metabolizing enzymes. Drug metabolizing CYP structures solved to date include 1A2 (24), 2A6 (25,26), 2C8 (27), 2C9 (28), 2D6 (29) and 3A4 (30,31) (Figure 1.2 with asterisks). The general folds are similar but active site topology varies dramatically between isoforms. This list leaves several drug-metabolizing CYPs whose structures are still unknown along with many other CYP proteins not involved in drug metabolism.

CYP2E1

The enzyme CYP2E1 was previously known as P450 "j" or ethanol-inducible P450. This enzyme was first purified from human liver (32) in 1987 and the human gene was characterized in 1988 (33). CYP2E1 is a microsomal CYP and is primarily hepatic but also can be found in extrahepatic tissues such as the lung, esophagus, and small intestine (34). It has been associated with toxicity through two major mechanisms. The causes of CYP2E1 associated toxicity are the generation of reactive metabolites, and generation of reactive oxygen species due to an uncoupled mechanism.

It has been proposed that CYP2E1 is more prone to reactive oxygen species production than other cytochrome P450 enzymes (35). In alcohol-induced liver disease (36), exposure to ethanol has been shown to induce CYP2E1 by stabilizing the protein and preventing its degradation. This increase in CYP2E1 protein content

has been associated with the cellular stress of alcohol-induced liver disease (36). Recently the propensity towards uncoupling leading to health problems has become more controversial. Studies utilizing *in vivo* levels of reactive oxygen species-mediated isoprostanes as measures of oxidative stress have shown no detectable alteration in oxidative stress with induction of CYP2E1 or in CYP2E1 knockout mice (37).

Two drugs CYP2E1 metabolizes to reactive metabolites are acetaminophen (38) and halothane (39,40). Acetaminophen is the most widely used analgesic in the U.S. (41) and one of the leading causes of fatal poisonings (38). Activation of acetaminophen by CYP2E1 into the strongly electrophilic N-acetyl-*p*-benzoquinone-imine (NAPQI) is the initial and major step responsible for acetaminophen hepatotoxicity (Scheme 1.2 **A**). Halothane is a volatile anesthetic whose use was discontinued due to severe hepatitis induced by CYP2E1 oxidation to trifluoroacetic acid chloride (TFAC) (Scheme 1.2 **B**). TFAC covalently modifies proteins to create an antigen that the body recognizes in later halothane exposures, resulting in severe hepatic necrosis known as halothane hepatitis (39).

In addition to acetomenophen and halothane, CYP2E1 is known to metabolize more than 70 low molecular weight (<100 g/mol) substrates (structures of representative substrates shown in Figure 1.4) including ~4% of clinically used drugs (19). Most of this catalysis acts to detoxify compounds but some leads to compound activation. CYP2E1 substrates include industrial solvents such as benzene, toluene, aniline, and halogenated solvents (42); alcohols such as ethanol (43), glycerol,

Scheme 1.2: Metabolic activation mechanism of acetaminophen (**A**) and halothane (**B**). (GSH = glutathione)

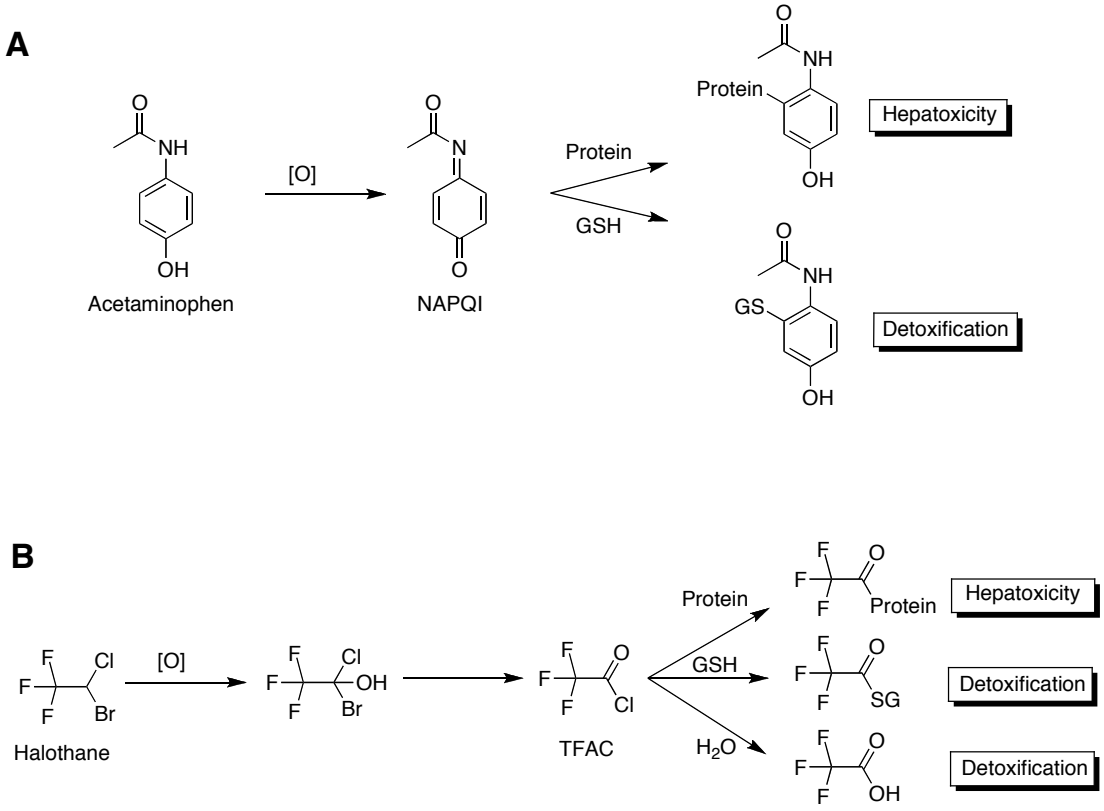
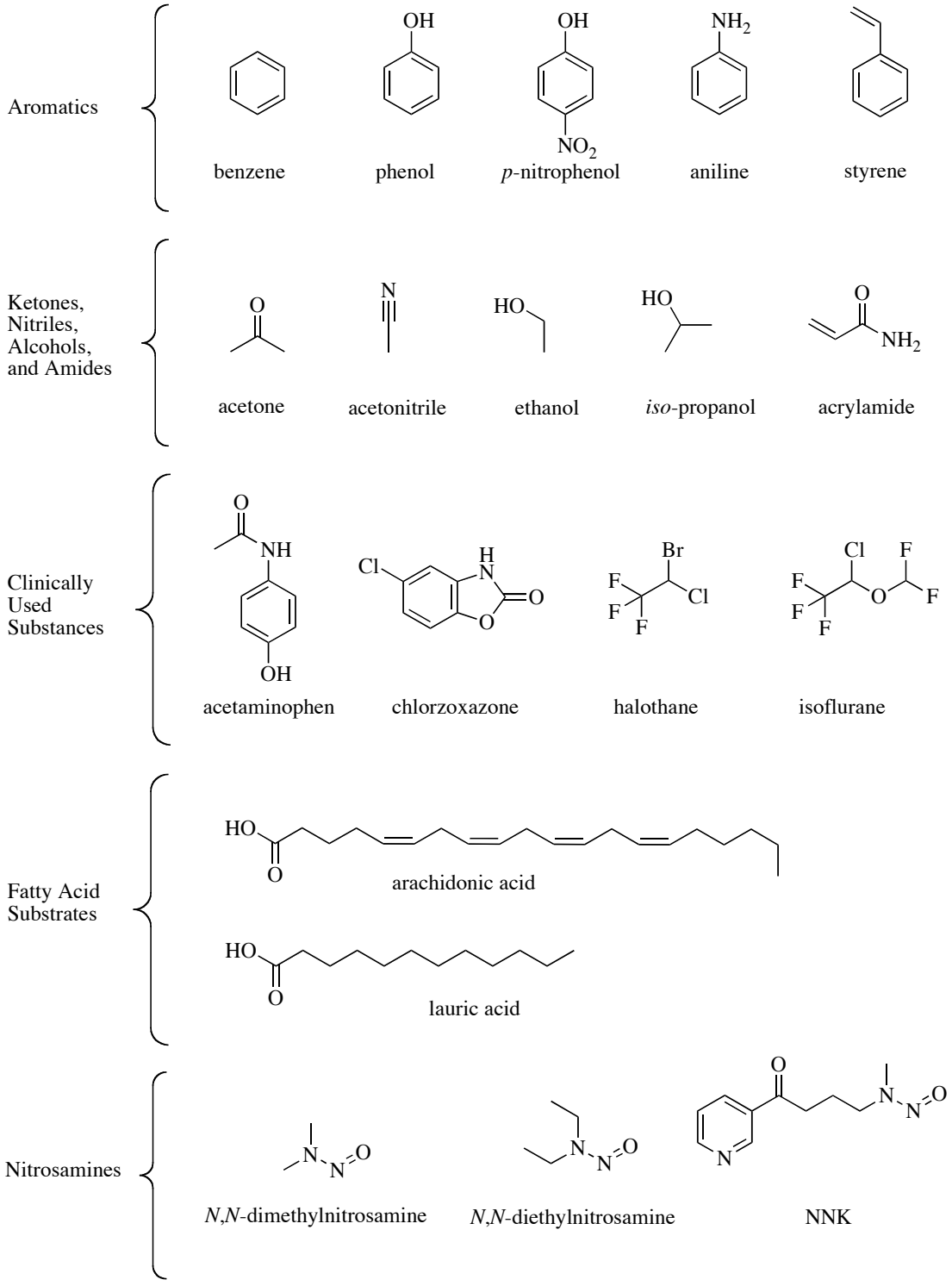


Figure 1.4: Representative CYP2E1 substrates.



phenol, and *p*-nitrophenol (44); and bicyclic heterocycles such as caffeine (45) and the muscle relaxant chlorzoxazone (46). However CYP2E1 is also known to metabolize endogenous fatty acids, including lipids associated with signaling mechanisms such as arachidonic acid (47) and epoxyeicosatrienoic acids (48). The role of CYP2E1 in the metabolism of these signaling molecules may be related to its regulation by disease states such as obesity (49) and diabetes in rat models (50).

Another class of compounds that CYP2E1 is known to metabolize are small nitrosamines. These compounds are activated to potent DNA alkylating agents. Some of the nitrosamines shown to be substrates are *N*,*N*-dimethylnitrosamine (51) and *N*,*N*-diethylnitrosamine (51) which are primarily metabolized by CYP2E1, *N*-nitrosopyrrolidine (52), *N*-nitroso-2,6-dimethylmorpholine (53), and to a lesser degree 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK) (54,55).

CYP2A13

CYP2A13 is specifically expressed in the respiratory tract (56) and has been shown to be the most efficient activator of the tobacco smoke-derived procarcinogen NNK (57,58). The gene for this enzyme was fully sequenced in 1995 (59). This enzyme is not known for the metabolism of drugs but has been associated with the activation of several pneumontoxins besides NNK including naphthalene (60), styrene (60), and aflatoxin B_1 (61). CYP2A13 has become an interesting drug target in the effort to prevent NNK metabolic activation and the subsequent lung toxicity.

Project Design

An important step in understanding the activation of NNK and in designing ways to inhibit its activation is to understand the structure of the enzyme responsible for its activation, CYP2A13. Previous work has focused on crystallizing and solving the structure of CYP2A13. The active site of CYP2A13 was occupied by an unknown ligand that had to be identified to complete the model of atoms in the experimentally determined electron density. Several pieces of evidence suggested that the ligand was indole, so initially we started evaluating existing colorometric assays, but they were not sufficiently sensitive or selective for indole over possible indole metabolites. A new method was developed that generates a chromophore and uses it as a handle for detection by UV/Vis detection before subsequent mass determination by MS/MS. This assay was utilized to detect unsubstituted indole in the protein mixture used to crystallize CYP2A13 allowing indole to be subsequently modeled into the CYP active site.

As previously discussed, CYP2E1 is known for both its metabolism of small molecular weight substrates as well as fatty acids. Several different approaches could be taken in order to understand the reasons for this substrate promiscuity. Through binding and inhibition studies, a model of the enzyme active site could be generated, but the identity of the specific residues responsible for substrate recognition would still be unknown. Through site directed mutagenesis the identity of specific residues involved could be identified but in some cases it is difficult to confirm that mutagenesis was not causing more drastic tertiary structural changes. Our approach

to understanding how CYP2E1 can bind and metabolize these seemingly divergent classes of compounds is through X-ray crystallography. This is a direct method for visualizing the residues involved in interacting and orienting substrates and inhibitors in the CYP2E1 active site. Though no experimental structures of CYP2E1 were known previously, I have determined CYP2E1 structures with a molecule from each of the classes: indazole representing small molecular weight compounds and ω-imidazolyl-decanoic acid representing fatty acids. These crystal structures have shown that subtle movements of active site amino acid side chains allow affinity for these two divergent compound classes.

Conclusion

The structural reasons for substrate selectivity among the isozymes discussed are extremely important. Understanding the active site topology of xenobiotic metabolizing enzymes can aid in predicting an isoform's metabolic profile. This structural knowledge also provides a rationale for designing drugs to either inhibit these CYPs to prevent their toxicological actions or to generate drugs that are not substrates of specific cytochrome P450 enzymes.

References

1. Garfinkel, D. (1958) *Arch Biochem Biophys* **77**(2), 493-509
2. Klingenberg, M. (1958) *Arch Biochem Biophys* **75**(2), 376-386
3. Omura, T., and Sato, R. (1962) *J Biol Chem* **237**, 1375-1376
4. Omura, T., and Sato, R. (1964) *J Biol Chem* **239**, 2370-2378
5. Omura, T., and Sato, R. (1964) *J Biol Chem* **239**, 2379-2385
6. Omura, T., and Sato, R. (1963) *Biochim Biophys Acta* **71**, 224-226
7. Beaune, P., Dansette, P., Flinois, J. P., Columelli, S., Mansuy, D., and Leroux, J. P. (1979) *Biochem Biophys Res Commun* **88**(3), 826-832
8. Wang, P., Mason, P. S., and Guengerich, F. P. (1980) *Arch Biochem Biophys* **199**(1), 206-219

9. Wang, P. P., Beaune, P., Kaminsky, L. S., Dannan, G. A., Kadlubar, F. F., Larrey, D., and Guengerich, F. P. (1983) *Biochemistry* **22**(23), 5375-5383
10. Bernhardt, R. (2006) *J Biotechnol* **124**(1), 128-145
11. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1986) *Biochemistry* **25**(18), 5314-5322
12. Das, A., Grinkova, Y. V., and Sligar, S. G. (2007) *J Am Chem Soc* **129**(45), 13778-13779
13. Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A., and Sligar, S. G. (2000) *Science* **287**(5458), 1615-1622
14. Li, H., Narasimhulu, S., Havran, L. M., Winkler, J. D., and Poulos, T. L. (1995) *Journal of the American Chemical Society* **117**(23), 6297-6299
15. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C., and Kraut, J. (1985) *J Biol Chem* **260**(30), 16122-16130
16. Joo, H., Lin, Z., and Arnold, F. H. (1999) *Nature* **399**(6737), 670-673
17. Kumar, S., Liu, H., and Halpert, J. R. (2006) *Drug Metab Dispos* **34**(12), 1958-1965
18. Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., and Nebert, D. W. (1996) *Pharmacogenetics* **6**(1), 1-42
19. Rendic, S., and Di Carlo, F. J. (1997) *Drug Metab Rev* **29**(1-2), 413-580
20. Guengerich, F. P., and Shimada, T. (1991) *Chem Res Toxicol* **4**(4), 391-407
21. Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000) *Mol Cell* **5**(1), 121-131
22. Scott, E. E., He, Y. A., Wester, M. R., White, M. A., Chin, C. C., Halpert, J. R., Johnson, E. F., and Stout, C. D. (2003) *Proc Natl Acad Sci U S A* **100**(23), 13196-13201
23. Scott, E. E., White, M. A., He, Y. A., Johnson, E. F., Stout, C. D., and Halpert, J. R. (2004) *J Biol Chem* **279**(26), 27294-27301
24. Sansen, S., Yano, J. K., Reynald, R. L., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2007) *J Biol Chem* **282**(19), 14348-14355
25. Yano, J. K., Denton, T. T., Cerny, M. A., Zhang, X., Johnson, E. F., and Cashman, J. R. (2006) *J Med Chem* **49**(24), 6987-7001
26. Yano, J. K., Hsu, M. H., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2005) *Nat Struct Mol Biol* **12**(9), 822-823
27. Schoch, G. A., Yano, J. K., Wester, M. R., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J Biol Chem* **279**(10), 9497-9503
28. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovic, D., and Jhoti, H. (2003) *Nature* **424**(6947), 464-468
29. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006) *J Biol Chem* **281**(11), 7614-7622

30. Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., and Jhoti, H. (2004) *Science* **305**(5684), 683-686
31. Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J Biol Chem* **279**(37), 38091-38094
32. Wrighton, S. A., Thomas, P. E., Ryan, D. E., and Levin, W. (1987) *Arch Biochem Biophys* **258**(1), 292-297
33. Umeno, M., McBride, O. W., Yang, C. S., Gelboin, H. V., and Gonzalez, F. J. (1988) *Biochemistry* **27**(25), 9006-9013
34. Ding, X., and Kaminsky, L. S. (2003) *Annu Rev Pharmacol Toxicol* **43**, 149-173
35. Bell, L. C., and Guengerich, F. P. (1997) *J. Biol. Chem.* **272**(47), 29643-29651
36. Lu, Y., and Cederbaum, A. I. (2008) *Free Radical Biol. Med.* **44**(5), 723-738
37. Dostalek, M., Hardy, K. D., Milne, G. L., Morrow, J. D., Chen, C., Gonzalez, F. J., Gu, J., Ding, X., Johnson, D. A., Johnson, J. A., Martin, M. V., and Guengerich, F. P. (2008) *J. Biol. Chem.* **283**(25), 17147-17157
38. Bronstein, A. C., Spyker, D. A., Cantilena, L. R., Jr., Green, J., Rumack, B. H., and Heard, S. E. (2007) *Clin. Toxicol. (Phila)* **45**(8), 815-917
39. Ray, D. C., and Drummond, G. B. (1991) *Br. J. Anaesth.* **67**(1), 84-99
40. Spracklin, D. K., Hankins, D. C., Fisher, J. M., Thummel, K. E., and Kharasch, E. D. (1997) *J Pharmacol Exp Ther* **281**(1), 400-411
41. Larson, A. M., Polson, J., Fontana, R. J., Davern, T. J., Lalani, E., Hynan, L. S., Reisch, J. S., Schiodt, F. V., Ostapowicz, G., Shakil, A. O., and Lee, W. M. (2005) *Hepatology* **42**(6), 1364-1372
42. Guengerich, F. P., Kim, D. H., and Iwasaki, M. (1991) *Chemical research in toxicology* **4**(2), 168-179
43. Koop, D. R., and Coon, M. J. (1986) *Alcohol Clin. Exp. Res.* **10**(6 Suppl), 44S-49S
44. Koop, D. R. (1986) *Mol. Pharmacol.* **29**(4), 399-404
45. Tassaneeyakul, W., Birkett, D. J., McManus, M. E., Tassaneeyakul, W., Veronese, M. E., Andersson, T., Tukey, R. H., and Miners, J. O. (1994) *Biochemical pharmacology* **47**(10), 1767-1776
46. Peter, R., Bocker, R., Beaune, P. H., Iwasaki, M., Guengerich, F. P., and Yang, C. S. (1990) *Chemical research in toxicology* **3**(6), 566-573
47. Laethem, R. M., Balazy, M., Falck, J. R., Laethem, C. L., and Koop, D. R. (1993) *J. Biol. Chem.* **268**(17), 12912-12918
48. Roy, U., Joshua, R., Stark, R. L., and Balazy, M. (2005) *Biochem. J.* **390**(Pt 3), 719-727
49. Salazar, D. E., Sorge, C. L., and Corcoran, G. B. (1988) *Biochemical and biophysical research communications* **157**(1), 315-320
50. Song, B. J., Matsunaga, T., Hardwick, J. P., Park, S. S., Veech, R. L., Yang, C. S., Gelboin, H. V., and Gonzalez, F. J. (1987) *Mol. Endocrinol.* **1**(8), 542-547

51. Yang, C. S., Yoo, J. S., Ishizaki, H., and Hong, J. Y. (1990) *Drug Metab Rev* **22**(2-3), 147-159
52. McCoy, G. D., and Koop, D. R. (1988) *Cancer Res* **48**(14), 3987-3992
53. Koop, D. R., and Coon, M. J. (1986) *Alcohol Clin Exp Res* **10**(6 Suppl), 44S-49S
54. Karamanakos, P. N., Trafalis, D. T., Geromichalos, G. D., Pappas, P., Harkitis, P., Konstandi, M., and Marselos, M. (2008) *Arch Toxicol*
55. Rodriguez-Antona, C., and Ingelman-Sundberg, M. (2006) *Oncogene* **25**(11), 1679-1691
56. Zhu, L. R., Thomas, P. E., Lu, G., Reuhl, K. R., Yang, G. Y., Wang, L. D., Wang, S. L., Yang, C. S., He, X. Y., and Hong, J. Y. (2006) *Drug Metab Dispos* **34**(10), 1672-1676
57. Wang, H., Tan, W., Hao, B., Miao, X., Zhou, G., He, F., and Lin, D. (2003) *Cancer Res* **63**(22), 8057-8061
58. von Weymarn, L. B., Chun, J. A., and Hollenberg, P. F. (2006) *Carcinogenesis* **27**(4), 782-790
59. Fernandez-Salguero, P., Hoffman, S. M., Cholerton, S., Mohrenweiser, H., Raunio, H., Rautio, A., Pelkonen, O., Huang, J. D., Evans, W. E., Idle, J. R., and et al. (1995) *Am J Hum Genet* **57**(3), 651-660
60. Fukami, T., Katoh, M., Yamazaki, H., Yokoi, T., and Nakajima, M. (2008) *Chemical Research in Toxicology* **21**(3), 720-725
61. He, X. Y., Tang, L., Wang, S. L., Cai, Q. S., Wang, J. S., and Hong, J. Y. (2006) *Int J Cancer* **118**(11), 2665-2671

Page left intentionally blank

Chapter 2.

Experimental Methods and Procedures

The general methods described in this chapter are used in the following chapters. The materials utilized were obtained from Fisher Scientific (Waltham, MA) unless otherwise noted.

CPY2E1 Expression and Purification

Protein Engineering

The coding region for CYP2E1dH with N-terminal truncations/modifications and a C-terminal 4xHis tag was generated as described for CYP2B enzymes (1) using the plasmid *pGEM 4-h2E1* as a template. This plasmid was kindly provided by Dr. M. Ingelman-Sundberg (Karolinska Institute). The resulting N-terminal and C-terminal amino acid sequences are MAKKTSSKGKLPPGP . . . PRSHHHH (non-native sequence underlined). The N-terminal modification aids in protein solubility by deleting the N-terminal transmembrane helix and incorporating charged residues, while the C-terminal His tag facilitates Ni-affinity purification. The details of generating the pKK2E1dH plasmid have been omitted because it was designed and generated by Drs. Emily E. Scott and Margit Spatzeragger before the initiation of this work (2). This truncated and 4xHis tagged protein, CYP2E1dH, will be abbreviated CYP2E1 throughout this thesis.

E. coli Transformation and Protein Expression

The expression process began with the transformation of the *E. coli* TOPP-3 (Stratagne, La Jolla, CA) cell line with the CYP2E1-containing plasmid, pKK2E1dH. TOPP-3 cells (50 µL) were incubated with the CYP2E1 plasmid (1 µL) for 30 minutes at 4 °C. The solution was heat shocked at 42 °C for 30 to 45 seconds and 1.7 µL of 10% (v/v in water) β-mercaptoethanol was immediately added. The cells were subjected to an outgrowth step consisting of the addition of 1 mL of super optimal catabolite (SOC) media and incubation at 37 °C for 1 hour with shaking at 250 rpm. After incubation the cells were pelleted by brief centrifugation and all but 100 µL of supernatant removed. The cells were resuspended in the remaining solution, spread on Luria-Bertani (LB) media agar plates with ampicillin at 50 µg/mL, and incubated at 37 °C overnight.

On the following day a single colony was inoculated into a 5 mL culture of LB media along with ampicillin (50 µg/mL) and tetracycline (25 µg/mL). The antibiotic treatment provided selection for TOPP-3 cells because of their genomic tetracycline resistance and for specific cells containing the pKK2E1dH plasmid carrying the ampicillin resistance gene. The 5 mL culture was incubated at 37 °C and shaken at 250 rpm for 8 hours. Overnight starter cultures, 2 X 200 mL LB with ampicillin (50 µg/mL) and tetracycline (25 µg/mL), were inoculated with 50 µL of the 5 mL culture, incubated at 37 °C, and shaken at 250 rpm overnight. Fifteen mL of this starter culture was used to inoculate each of 18 1L flasks each containing 250 mL of terrific broth (TB) with ampicillin (50 µg/mL). These flasks were grown to an

$OD_{600\ nm}$ of 1.0-1.5 at 37 °C and shaken at 250 rpm (~2 hours). The temperature was decreased to 30 °C, shaking decreased to 190 rpm, and the media supplemented with δ-aminolevulinic acid (20 mg/flask) as a heme precursor and isopropyl β-D-1-thiogalactopyranoside (60 mg/flask) to induce protein expression by binding the Lac repressor. Cells were grown for two days and harvested by centrifugation at 6300 x g. The supernatant was discarded and pellets were stored at -80 °C until use.

Protein Purification and Ligand Complexation

E. coli cells expressing CYP2E1 were resuspended in 200 mL of 20 mM potassium phosphate buffer pH 7.4 containing 20% glycerol and lysozyme added to 0.3 mg/mL and stirred at 4 °C for 30 minutes. Slowly 200 mL of water at 4 °C was added and the solution stirred at 4 °C for 10 minutes. The spheroplasts were pelleted by centrifugation at 8800 x g for 15 minutes. The supernatant was discarded and pellet resuspended in 100 mL of 100 mM potassium phosphate buffer pH 7.4 containing 20% glycerol and 1.0 M NaCl. The solution was homogenized with a tissue homogenizer and separated into three aliquots for sonication. The solutions were each sonicated three times on ice for 30 seconds (Fisher Scientific, Sonic Dismembrator, Model 100). Cellular debris was removed by centrifugation at 8800 x g for 15 minutes. The supernatant was retained and Cymal-5 (Anatrace, Maumee, OH) was added to 4.8 mM and stirred at 4 °C for 60 min to solublize CYP2E1. The solution was ultracentrifuged at 80,000 X g for 60 min to clarify it before the subsequent chromatographic steps. The resulting supernatant was applied to NiNTA fastflow resin (Qiagen, Valencia, CA) and washed with 100 mM potassium phosphate

buffer pH 7.4 containing 20% glycerol, 300 mM NaCl, and 4.8 mM Cymal-5 to wash out unbound sample. The column was then washed with 100 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 200 mM NaCl, 15 mM imidazole, and 4.8 mM Cymal-5 to remove nonspecific proteins. CYP2E1 was eluted with 50 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 100 mM NaCl, 180 mM imidazole, 4.8 mM Cymal-5, and 10 mM EDTA. CYP2E1 fractions were pooled and diluted 5-fold with 5 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 5 mM EDTA, and 4.8 mM Cymal-5 to lower the ionic strength and facilitate binding to ion exchange resin. This solution was applied to a carboxymethyl cellulose column (GE Healthcare, Uppsala, Sweden), washed with the dilution buffer without detergent to remove excess Cymal-5 from the protein, and CYP2E1 eluted with 100 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 500 mM NaCl, and 1 mM EDTA. CYP2E1 fractions were concentrated to ~1 mL using Amicon Ultra-4 centrifugal ultra filtration devices (Millipore, Billerica, MA, 50,000 molecular weight cut off), loaded onto a Superdex 200 16/60 gel-filtration column (GE Healthcare, Uppsala, Sweden) and eluted with 100 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 500 mM NaCl, and 1 mM EDTA. The final CYP2E1 fractions eluted from gel-filtration were pooled, concentrated using Amicon Ultra-4 centrifugal ultra filtration devices (Millipore, Billerica, MA, 50,000 molecular weight cut off), and the buffer exchanged through concentrating to ~100 μL followed by diluting solution with ~4 mL of 120 mM potassium phosphate pH 7.4, 0.5 M sucrose, and 1 mM EDTA containing either 5

mM indazole or 25 mM ω-imidazolyl-decanoic acid. This concentrating and dilution process was repeated 3-5 times to ensure buffer exchange. All columns were monitored by UV/Vis absorbance at the wavelengths of 280 nm, 393 nm, and 424 nm.

Spectroscopic Techniques

Absolute Spectra (3)

To analyze protein concentration and purity during purifications as well as the final product, an absolute spectrum was used. The absorbance spectrum was traditionally measured from 750 nm to 250 nm. The protein spectra were background –subtracted with a buffer sample. The concentration of the protein was determined using the Beer-Lambert law with the extinction coefficient of 0.1 mL / (nmol ∗ cm) at 424 nm. The relative purity was judged by comparing the Soret maximum (major heme absorption band occurring between 390 nm and 424 nm depending on ligand bound state) with the absorbance at 280 nm due tryptophan absorbance from both CYP2E1 and other proteins.

CO Difference Assays (3)

Another way that CYP was quantified is through the spectrophotometric change with carbonmonoxide binding. This method was used to quantify protein for the purpose of spectral binding assays. A protein sample was prepared and a few grains of sodium dithionite (Sigma Aldrich, St. Louis, MO) added as a reducing agent (4). The difference spectrum was acquired by the subtraction of the reduced protein

absorbance from the absorbance of reduced protein after exposure to 50 bubbles of CO (Air Liquide, Houston, TX). The concentration of the protein was determined using the Beer-Lambert law with the extinction coefficient of 0.091 mL / (nmol ∗ cm) at 450 nm.

Spectral Binding Assays (3)

Ligand binding was determined by monitoring the change in the Soret peak upon ligand binding. Typically in the resting state the CYP Soret is at ~414 nm corresponding to six coordinate heme iron with water bound at the sixth position. Upon binding of a ligand, a water is displaced from the heme iron. For a "Type I" ligand, the water is displaced and the ligand does not coordinate to the heme iron yielding five coordinate heme iron and shifting the Soret to ~390 nm. If the ligand displaces the water and binds directly to the heme iron (coordinate covalently typically observed with a basic nitrogen) the Soret shifts to ~424 nm, which is indicative of a "Type II" ligand. Experimentally, the following procedure was utilized to measure these spectral binding constants. Two cuvettes were charged with 800 µL of CYP at 1 µM diluted in 100 mM KPi pH 7.4, 20% glycerol. One cuvette becomes the vessel for measuring ligand effects on absorbance while the other was used to correct for solvent effects. An aliquot of ligand dissolved in vehicle is added to the sample cuvette along with an equal volume of vehicle to the reference cuvette and both are mixed. The solutions are allowed to equilibrate for 5 minutes and the absorbance measured from 300 nm to 500 nm. As this process was repeated, a peak and a trough form and the process was continued until there was no further change in

the peak to trough difference. The peak to trough difference (ΔA) was plotted verse ligand concentration using Graphpad Prism 4 (Graphpad Software, San Diego, CA). To determine K_D (the spectrally observed dissociation constant) and ΔA_{max} (the maximal peak to trough absorbance change) the data were fit using nonlinear least-squares regression to the following equation: $\Delta A = \frac{\Delta A_{max} \times [ligand]}{K_D + [ligand]}$.

Crystallographic Techniques and Structure Analysis

Hanging-drop Vapor Diffusion

All crystallization protocols that will be described in subsequent chapters were performed using the hanging drop vapor diffusion crystallization method. This method utilized VDX plates (Hampton Research, HR3-142, Aliso Viejo, CA) greased with high vacuum grease around each well lip of the 24 well plate. Each well was loaded with 750 µL of well solution. In the center of a siliconized glass cover slip (Hampton Research, HR3-239, Aliso Viejo, CA) 1 µL of protein solution was placed. To this drop, 1 µL of well solution was directly added but not mixed. The slip is inverted over the well and pressed against vacuum grease for an airtight seal. The crystallization drops were allowed to equilibrate at 20 °C and were monitored for crystal growth. Specific materials and methods utilized for the individual crystal structures including crystallization conditions, cryoprotectants, data collection, data processing, and phasing are discussed in the corresponding chapters.

VOIDOO and CAVER

Probe-occupied voids were calculated using VOIDOO (5) with a probe radius of 1.4 Å and a grid mesh of 0.3 Å unless otherwise specified. CAVER (6) calculations were performed utilizing a PyMOL plugin and a grid mesh of 0.8 Å.

References

1. Scott, E. E., Spatzenegger, M., and Halpert, J. R. (2001) *Arch Biochem Biophys* **395**(1), 57-68
2. Spatzenegger, M., Liu, H., Wang, Q., Debarber, A., Koop, D. R., and Halpert, J. R. (2003) *J Pharmacol Exp Ther* **304**(1), 477-487
3. Schenkman, J. B., and Jansson, I. (2006) *Methods Mol Biol* **320**, 11-18
4. Omura, T., and Sato, R. (1962) *J Biol Chem* **237**, 1375-1376
5. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**(Pt 2), 178-185
6. Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006) *BMC Bioinf.* **7**, 316

Chapter 3.

Identification of an Unknown Ligand Bound in the CYP2A13 Crystal Structure: Teaching an Old Assay New Tricks

The experiments described in this chapter have been previously published in: Porubsky, P. R., Scott, E. E., and Williams, T. D. (2008) *Arch Biochem Biophys* **475**(1), 14-17. Additionally, this assay was used to identify the ligand bound to CYP2A13, leading to the publication of its structure in: Smith, B. D., Sanders, J. L., Porubsky, P. R., Lushington, G. H., Stout, C. D., and Scott, E. E. (2007) *J. Biol. Chem.* **282**(23), 17306-17313.

Introduction

The lung CYP2A13 activates the potent tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) into two possible DNA alkylators (1-3). In an effort to more fully understand this enzyme's structure and function, CYP2A13 has been crystallized and X-ray diffraction experiments performed. These experiments illuminated the structure of this enzyme, but an unknown ligand was present in the enzyme active site (4). This ligand was either co-purified with the protein or added adventitiously. The three following pieces of data suggested that the identity of the unknown ligand was indole or an indole metabolite. First, a 94% homologous enzyme, CYP2A6, is known to catalyze the oxidation of indole to several different oxidized products and pigments (5). Some of these pigments were also visible during the *E. coli* expression of CYP2A13. Second,

indole is present in the *E. coli* expression system due to the inherent tryptophanase activity of the bacteria. Third, the electron density of the ligand in the crystal structure appeared to be for a planar bicyclic compound but the density covers enough area that ring substitutions could be possible ($2|F_o|-|F_c|$ electron density map at 1.0 σ shown in blue mesh in Figure 3.1). In order to determine the identity of this ligand and especially to differentiate between indole and indole metabolites that could have been produced by CYP2A13 metabolism of indole, we developed a sensitive HPLC-UV/Vis-MS/MS assay (4).

Several different aldehyde-containing reagents have been used in microbiology as tools to differentiate certain genera and species of bacteria. A positive colorimetric result demonstrates that the specific bacterium produces tryptophanase, an indole lyase. Some of the reagents used for indole detection are *p*-dimethylaminobenzaldehyde (utilized in both the Ehrlich indole reagent and the Kovacs indole reagent) and *p*-dimethylaminocinnamaldehyde (DMACA). DMACA is both more rapid and long lasting, yielding results that are easier to interpret in bacterial assays (6). Two advantages for our purposes were the superior sensitivity of the DMACA reagent and its ability to detect indole derivatives (6,7). In fact, DMACA has previously been utilized for the quantitative measurement of indole-3-methanol, also known as indole-3-carbinol, in purified plant protein samples (8). A modified method of indole conjugate formation, discussed herein, has been developed to distinguish the identities of indole and indole-like compounds by spectrophotometric and HPLC-UV/Vis-MS/MS methods with increased sensitivity

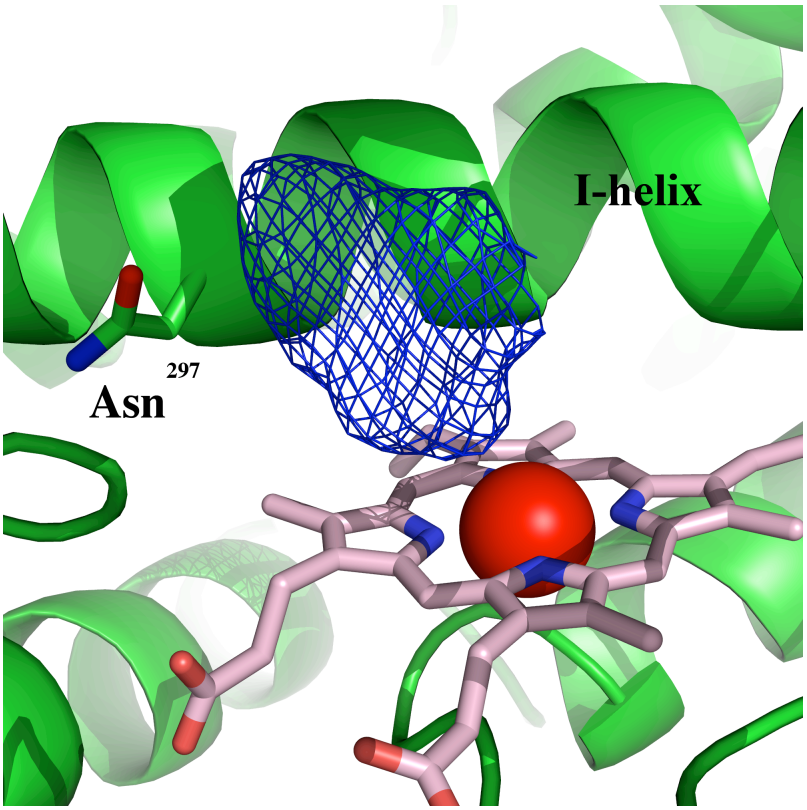


Figure 3.1: CYP2A13 active site with $2|F_o|-|F_c|$ electron density map contoured at 1.0 σ (blue mesh). The density is planar and with sufficient bulk for a bicyclic compound.

over previously reported techniques. The data from a set of known compounds was used as standards to identify the unknown ligand in the purified protein solution used to generate the CYP2A13/ligand X-ray crystallographic structure (4).

Methods

Materials

Compounds 6-hydroxyindole, 5-hydroxyindole, 4-hydroxyindole, 3-hydroxyindole, isatin and tryptophan were obtained from Acros (Geel, Belgium), indole from Aldrich (St. Louis, MO), oxindole from Maybridge (Trevillett, UK) and *p*-dimethylaminocinnamaldehyde (DMACA) from Fluka (Japan). Purified CYP2A13 previously prepared for crystallographic purposes was obtained from Brian Smith (Scott Lab, Medicinal Chemistry) for analysis of unknown ligand. All other materials were obtained from Fisher Scientific.

Spot Tests and λ_{max} Determinations

In order to determine which indole compounds should be analyzed, initial spot tests were performed by saturating filter paper with DMACA solution (DMACA solution: 0.117 g of DMACA, 39 mL of ethanol, 5 mL of concentrated aqueous HCl and diluted to 50mL with water (6)), then spotting a solution containing indole or an indole derivative on the DMACA saturated filter paper with a capillary tube. The compounds that turned blue like the protein/ligand derivative were analyze further. To quantify these observations, this derivatization was modified to a spectrophotometric assay. In each coupling reaction 500 μL of indole containing solution (concentration between 10 ng/mL and 10 μg/mL) was combined with 250 μL

of DMACA solution in a cuvette. The absorbance was then measured from 400 to 750 nm. When purified protein samples were used, the protein (500 μL of 65 μM P450) was combined with DMACA reagent. This precipitated the protein and thus the mixture was centrifuged at 10,600 x g for 2 minutes to clarify the solution before acquisition of a spectrum. Controls were performed with indoles dissolved in the same buffer system as the protein to rule out complicating matrix effects on the λ_{max} values of conjugates. Consistent λ_{max} values were observed over the concentration ranges tested (±1 nm) and in the protein buffer matrix.

LC-UV/Vis-MS/MS Analysis of Derivatives

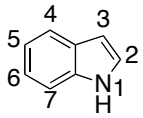
Identical conjugation conditions were used for LC-UV/Vis-MS/MS detection with indole concentrations of 0.2 pmol/μL in standards and 65 μM protein for ligand analysis. Separation was accomplished on 50 X 4.6 cm Alltech Altima C18 column with 3.0 μm particle size running at 0.3 mL/min. HPLC was performed with the mobile phases (Solvent A) 99% H_2O, 1% methanol, 0.06% formic acid, 10 mM ammonium formate and (Solvent B) 99% methanol, 1% H_2O, 0.06% formic acid, 10 mM ammonium formate. The solvent gradient was run to 1% B by 1 min (Δ% B of 1 %/min), then to 50% B by 2 min (Δ% B of 49 %/min), and then to 80% B by 15 min (Δ% B of 2.3 %/min). The UV/Vis detector was set at 625 nm for all LC-UV/Vis-MS/MS experiments. A Waters 2695 chromatograph was used to develop gradients and present peaks to a Micromass Quattro Ultima "triple" quadrupole mass spectrometer with an electrospray source.

Results and Discussion

The reaction between the unknown ligand from the CYP2A13 protein mixture and DMACA generated a conjugate that turned blue using the spot test and absorbed at 625 nm. A set of indoles was initially screened by a spot test on filter paper saturated with DMACA solution. The compounds yielding a blue result from the spot test became candidates for the unknown ligand. Indoles were tested at concentrations between 10 μg/mL and 10 ng/mL (85 μM to 85 nM for indole, concentrations vary with molecular weight with respect to the other indoles).

From the initial UV/Vis analysis, most of the oxidized indoles could be ruled out as the CYP2A13 ligand (Table 3.1). The λ_{max} of the UV/Vis data suggested that the ligand (derivatized λ_{max} = 625 nm) was consistent with corresponding results from either indole (derivatized λ_{max} = 623 nm) or 5-hydroxyindole (derivatized λ_{max} = 627 nm). Thus, the purely spectrophotometric method demonstrated that a highly absorbing conjugate of the CYP2A13 ligand could be formed and was successful at narrowing the identity of the CYP2A13 ligand. This characteristic was subsequently used for identification of the ligand conjugate on the HPLC phase of HPLC tandem MS.

Table 3.1: Comparison of results gathered from indoles tested (ND: not determined)

 (Indole Numbering System)	Spot Test	λ_{max} (nm) of derivative	Retention Time (min)	MW of derivative (experimental $M+H^+$)
indole	blue	623	8.7	275
3-hydroxyindole	purple	ND	ND	ND
4-hydroxyindole	blue	652	8.0	291
5-hydroxyindole	blue	627	7.3	291
6-hydroxyindole	blue	632	7.2	291
tryptophan	red	578, 480	ND	ND
oxindole (indolin-2-one)	red	ND	ND	ND
isatin (1H-indole-2,3-dione)	orange	ND	ND	ND
unknown	blue	625	8.6	275

From previous literature and the results of the spot test, we hypothesized that the reaction that resulted in the conjugates absorbing in the 600-700 nm wavelength region was a different regiochemical reaction than the reaction of indole-3-methanol reported by Muñoz and coworkers (8). Muñoz reported the generation of an azafulvenium salt from the reaction of indole-3-methanol with DMACA (Figure 3.2 **A**). This reaction occurred at the 2 position on the indole nucleus because the 3 position was blocked by the substituent. It has been reported that in the reaction of indole and retinal, the 3 position of indole is the most nucleophilic when not blocked (9). Likewise, in the Vilsmeir-Haack formylation of indole, selective electrophilic addition to indole is made at the 3 position (10). The compounds from our set of indoles that react with DMACA most rapidly and give spectrophotometric responses at the lowest concentrations are those with the unhindered 3 position. It may be that the free 3 position of the studied indoles was the reason for the increased sensitivity (proposed reaction in Figure 3.2 **B**).

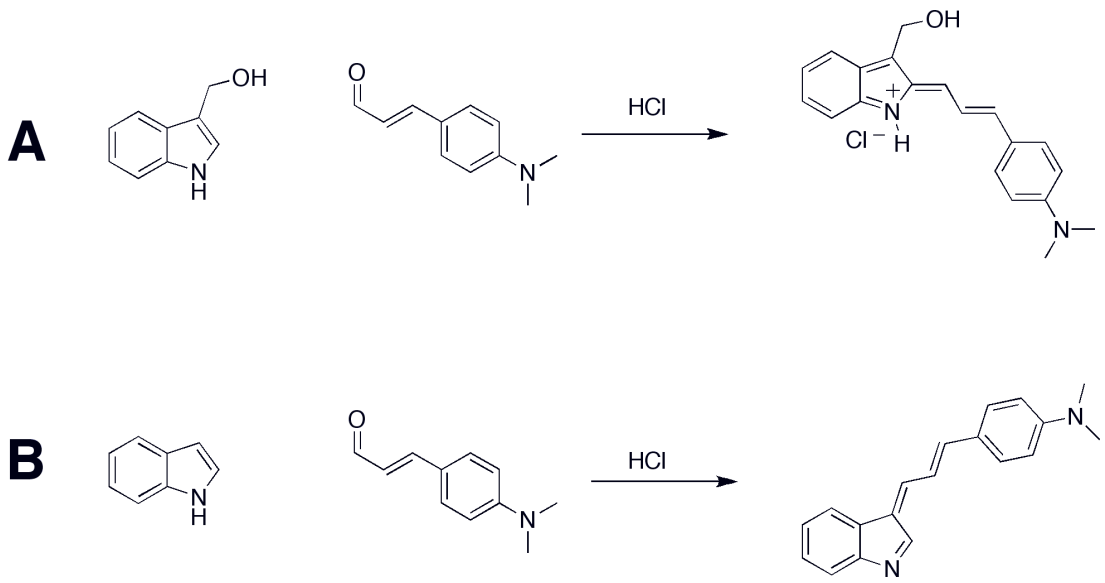


Figure 3.2: Proposed reaction schemes of indole-3-methanol (as proposed in (8), **A**) and unsubstituted indole (**B**) with *p*-dimethylaminocinnamaldehyde to form highly conjugated chromophores.

Standards of DMACA conjugates for 4-hydroxyindole, 5-hydroxyindole, 6-hydroxyindole, and indole were tested to determine retention times and separability with injection concentrations of indoles at 0.2 pmol/μL. UV/Vis detection at 625nm (λ_{max} of conjugate in protein mixture) of 4-hydroxyindole and 6-hydroxyindole DMACA conjugates were significantly lower than that of the other two conjugates due to their lower molar absorptivity at the detection wavelength. It was also observed that the sample analysis was optimal when analyzed the same day. Samples that were exposed to ambient conditions overnight showed significant decomposition of the DMACA conjugate. The predicted products of DMACA and indole were detected as MH^+ coinciding with the 625 nm HPLC-UV/Vis peak. In subsequent injections, product ion scans reveal a common fragment at m/z = 158 (Figure 3.3) in all derivatives tested. Utilizing multiple reaction monitoring (MRM) targeted MS/MS detection with two transitions of 274 and 291 to 158, the HPLC chromatographic resolution was preserved with a fast detection cycle line (< 1 s). Figure 3.3 illustrates the MRM (**A**) and HPLC-UV/Vis detection (**B**) of a standard mixture of DMACA conjugates of 5-hydroxyindole and indole. Adjacent are similar traces of the derivative from the CYP2A13/ligand mixture (**C** and **D**, respectively). The MS/MS scan of conjugated indole in the standard mixture is presented in **E**. DMACA conjugation and extraction of the protein/ligand sample presented an HPLC peak with a product ion spectrum identical to the standard indole product ion spectrum shown in **E**.

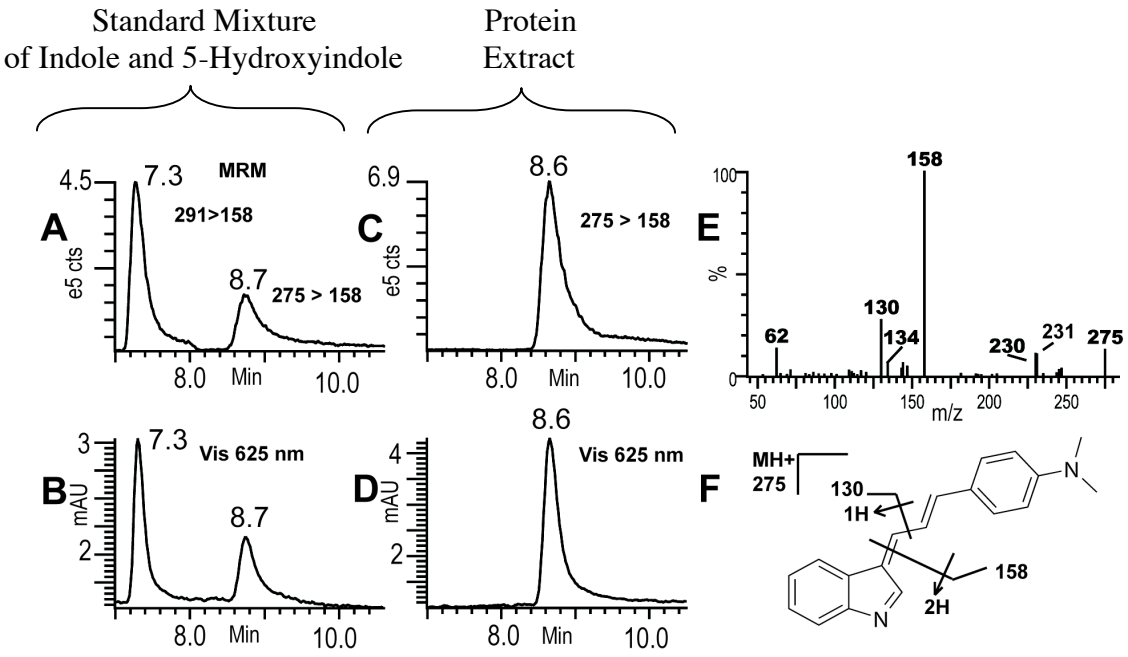


Figure 3.3: HPLC-Vis-MS/MS traces from indole standards and protein extract. A mixture of indole and 5-hydroxyindole standard conjugates (200 nM each) detected with MS/MS transition of (**A**) m/z 291 to 158 (from 5-hydroxyindole conjugate) and 274 to 158 (from indole conjugate), and **B** 625 nm HPLC trace. Spectra corresponding to **A** and **B** but from protein extract are illustrated in **C** and **D**, respectively. The MRM transitions were chosen from product ions observed in standard indole derivative product ion spectrum **E**. The proposed structure and identity of major ions of conjugated indole are illustrated in **F**. (cts = counts, mAU = milli-absorbance unis)

The proposed structure assignment (**F**) is supported by the common ion at m/z = 158 between all derivatives and that m/z = 130 for the indole derivative shifts to m/z = 146 in hydroxyindole samples. The HPLC-UV/Vis-MS/MS experiment positively confirmed the presence of indole in the purified protein/ligand mixture.

The information generated by this assay provided the rationale for modeling unsubstituted indole into the CYP2A13 active site (Figure 3.4). When indole is modeled in a single orientation so that it directly hydrogen bonds with Asn^{297} (yellow indole in Figure 3.4), a significant amount of electron density remains above the 5-position of indole. The analytical examination of the protein samples showed no detectable 5-hydroxyindole in the mixture, so it was hypothesized that unsubstituted indole may also occupy a second orientation in the active site causing this density. The alternate orientation involves indole rotating ~80° from its previous location. It was then modeled such that a water molecule bridges between the indole nitrogen and Asn^{297} (teal indole and small red sphere in Figure 3.4). The final model of the CYP2A13 structure refined to have these two orientations with near equal occupancy with small deviations between different molecules in the asymmetric unit of the crystalline lattice.

Conclusions:

While the colorimetric spot test for indole production has been in common use in microbiology for years, to our knowledge this is the first account utilizing the strongly absorbing chromophore generated in the conjugation as a handle for LC-

UV/Vis detection with subsequent MS/MS for unambiguous identification of indoles. The derivative, while relatively short lived (1/2 day), greatly improved the UV/Vis and HPLC detectability by moving the absorbance into the visible region and out of small aromatic absorbing region, increasing HPLC selectivity, and enhanced MS efficiency for indoles with unhindered 3 positions. This method of derivatization and detection of indoles was successful in confirming the presence of unsubstituted indole in the crystallization solution and allowed indole to be modeled into the CYP2A13 active site. In the future, the utility of this assay can be used to identify indole metabolites generated by CYP2A13.

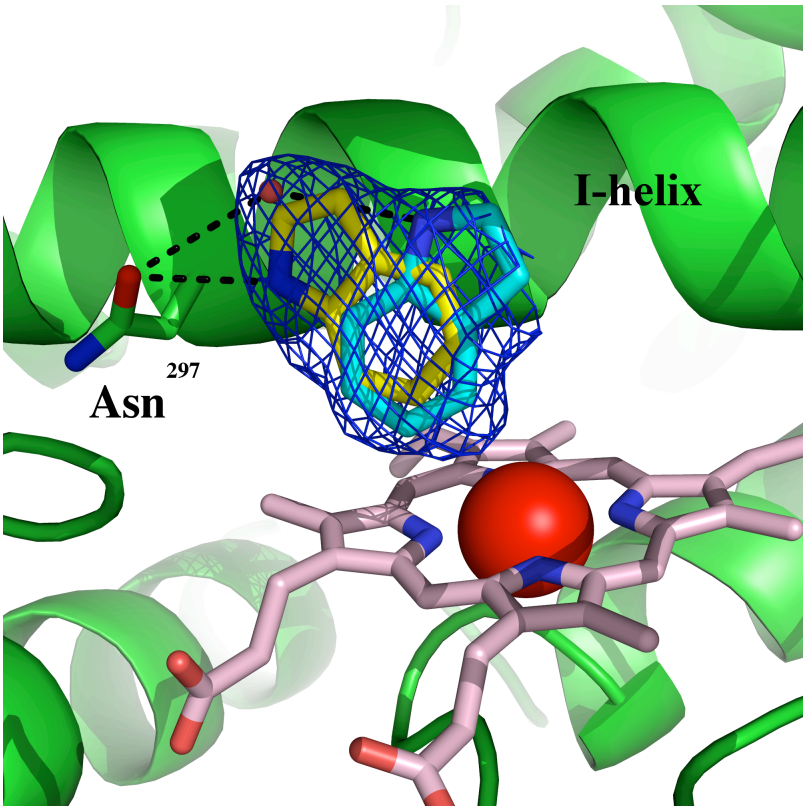


Figure 3.4: CYP2A13 active site with $2|F_o|-|F_c|$ electron density map at 1.0 σ (blue mesh) and indole modeled in two alternate conformations. Each alternate conformation has a different hydrogen bonding opportunity. The indole in conformation A (yellow indole) is within direct hydrogen bonding distance to Asn^{297}. Alternately, indole in conformation B (teal indole) is proposed to interact with Asn^{297} through a modeled water molecule (small red sphere).

References

1. He, X.-Y., Shen, J., Ding, X., Lu, A. Y. H., and Hong, J.-Y. (2004) *Drug Metab Dispos* **32**(12), 1516-1521
2. Jalas, J. R., Ding, X., and Murphy, S. E. (2003) *Drug Metab Dispos* **31**(10), 1199-1202
3. Su, T., Bao, Z., Zhang, Q. Y., Smith, T. J., Hong, J. Y., and Ding, X. (2000) *Cancer Res* **60**(18), 5074-5079
4. Smith, B. D., Sanders, J. L., Porubsky, P. R., Lushington, G. H., Stout, C. D., and Scott, E. E. (2007) *J. Biol. Chem.* **282**(23), 17306-17313
5. Gillam, E. M., Notley, L. M., Cai, H., De Voss, J. J., and Guengerich, F. P. (2000) *Biochemistry* **39**(45), 13817-13824
6. Lombard, G. L., and Dowell, V. R., Jr. (1983) *J Clin Microbiol* **18**(3), 609-613
7. Miller, J. M., and Wright, J. W. (1982) *J Clin Microbiol* **15**(4), 589-592
8. GarcÌa-Florenciano, E., Ros Barcelo, A., Sabater, F., and Munoz, R. (1989) *Analytical Biochemistry* **183**(1), 172-176
9. Johnston, D. S., Tyman, J. H. P., and Chapman, D. (1987) *Tetrahedron* **43**(18), 4177-4184
10. James, P. N., and Snyder, H. R. (1959) *Organic Syntheses* **39**, 30

Page left intentionally blank

Chapter 4.

Structure of Cytochrome P450 2E1 in Complex with Indazole

The experiments described in this chapter have been previously published as: Porubsky, P. R., Meneely, K. M., and Scott, E. E. (2008) *J Biol Chem* **283**(48), 33698-33707. The atomic coordinates and structure factors of the structure described in this chapter have been deposited in the protein data bank (code 3E6I).

Introduction

Cytochrome P450 (P450) is a superfamily of enzymes involved in monooxygenation of both endogenous and exogenous substrates. A subset of these enzymes, including cytochrome P450 2E1 (CYP2E1), are known for their role in the clearance of drugs and other xenobiotics by introducing or unmasking sites for subsequent conjugation and elimination from the body. From a biochemical standpoint these enzymes are fascinating for the diversity of substrates each xenobiotic-metabolizing P450 can metabolize, yet often with exquisite selectivity in the metabolites generated. Unfortunately, in the process some P450-mediated metabolism can produce toxic or carcinogenic products. Among cytochrome P450 enzymes, CYP2E1 is particularly notable for this ability and the resulting toxicity (1). It is primarily hepatic comprising over 50% of the hepatic cytochrome P450 mRNA (2) and 7% of the hepatic cytochrome P450 protein (3). However, CYP2E1 is also expressed at lower levels in a variety of extrahepatic tissues (4), where it is thought to play a role in the metabolism of important endogenous molecules. CYP2E1 levels and toxicity varies markedly in response to alcohol consumption (5), diabetes (6),

obesity (7), and fasting (8). Thus, the action of CYP2E1 can have significant influence on human health and drug metabolism.

CYP2E1 has been connected with liver toxicity through two mechanisms: by the activation of substrates into reactive metabolites and potentially by generation of reactive oxygen species (ROS). As described in Chapter 1, when substrate binding, oxygen binding, and electron delivery are not closely coordinated during catalysis, diatomic oxygen can be converted to superoxide or hydrogen peroxide and released instead of being used to monooxygenate the substrate. These ROS have been implicated in a range of damaging events including lipid peroxidation, protein oxidation, and DNA oxidation. CYP2E1 is considerably more prone to ROS production than other cytochrome P450 enzymes (9). For example, evidence shows that ethanol consumption substantially increases the levels of CYP2E1 protein and the resulting ROS have been suggested as the primary contributors to the negative physiological consequences of alcohol-induced liver disease (10). However studies with induction of CYP2E1 or CYP2E1 knockout mice have not detected alterations in the *in vivo* levels of ROS-mediated isoprostanes, a measure of oxidative stress (11).

Two well-studied drugs that are converted to reactive metabolites by CYP2E1 are acetaminophen (12,13) and halothane (14). Acetaminophen is the most widely used analgesic in the U.S. (15) and one of the leading causes of fatal poisonings (16). Activation of acetaminophen by CYP2E1 into the strongly electrophilic N-acetyl-*p*-benzoquinone-imine (NAPQI) is the initial and major step responsible for acetaminophen hepatotoxicity. Halothane is a volatile anesthetic whose use was

discontinued due to severe hepatitis induced by CYP2E1 oxidation to trifluoroacetic acid chloride (TFAC). TFAC covalently modifies proteins to create an antigen that the body can recognize in later halothane exposures, resulting in severe hepatic necrosis known as halothane hepatitis (17).

Acetaminophen and halothane are just two clinical representatives of the >70 low molecular weight (<100 g/mol) substrates that CYP2E1 is known to metabolize. These substrates include industrial solvents such as benzene, toluene, aniline, and halogenated solvents (18); alcohols such as ethanol (19), glycerol, phenol, and *p*-nitrophenol (20); and bicyclic heterocycles such as caffeine (21) and the muscle relaxant chlorzoxazone (22). However CYP2E1 is also known to metabolize endogenous fatty acids, including lipids associated with signaling mechanisms such as arachidonic acid (23) and epoxyeicosatrienoic acids (24). The role of CYP2E1 in the metabolism of these signaling molecules may be related to its regulation by disease states such as obesity (25) and diabetes (26) in rat models.

Although structures of several other important drug-metabolizing cytochromes P450 have been solved in recent years and a number of CYP2E1 models have been proposed, there are no experimental structures for CYP2E1. Cytochromes P450 have similar overall tertiary structures, but differ substantially in the size and shape of their active sites and the presence and topology of channels allowing substrate entry and metabolite egress, factors that in turn determine substrate selectivity in this enzyme superfamily. In order to better understand the ability of CYP2E1 to efficiently bind and metabolize both small molecular weight compounds

and fatty acids, the structure of this enzyme was determined by X-ray crystallography bound with the small molecular weight inhibitor indazole. Comparison of CYP2E1 structures to those of other human P450 isoforms provides insight into the functional similarities and differences among the human cytochrome P450 enzymes.

Methods

Protein Engineering, Expression, Purification, and Ligand Complexation

See Methods (Chapter 2) for the description of protein engineering, expression, purification, and formation of the ligand complex.

Protein Crystallization, Data Collection, and Structure Determination

Crystals were grown by hanging drop vapor diffusion as described in Methods (Chapter 2). The CYP2E1/ indazole complex (50 mg/mL CYP2E1, 5 mM indazole) was equilibrated against 0.1 M NaHEPES pH 7.5, 11% *iso*-propanol, and 6% PEG 2000 MME. Crystals were immersed in 0.1 M NaHEPES pH 7.5, 10% *iso*-propanol, 1.4 M sucrose as a cryoprotectant before being flash cooled in liquid nitrogen for data collection. A single native data set was collected for each complex on beamline 9-2 at the Stanford Synchrotron Radiation Laboratory (Stanford, CA) with a wavelength of 0.979 Å at 100 K. Data were processed using Mosflm (27) and Scala (28). The CYP2E1/ indazole structure was solved by molecular replacement using the program PHASER (29) and a search model consisting of CYP2A13 (PDB 2P85) with non-identical side chains truncated at the last common atom. Iterative model building and refinement were performed using COOT (30) and REFMAC (31).

Structure validation was performed using WHAT-IF (32) and PROCHECK (33). For the indazole bound CYP2E1 structure the Ramachandran plot showed 89.4% of the amino acid residues in the most favorable region, 9.3% in the additional allowed region, 0.8% in the generously allowed region, and 0.5% in the disallowed region.

Structure Analysis

Voids and exit channels were probed using VOIDOO (34) and CAVER (35) as described in Methods (Chapter 2). Secondary structure was determined using DSSP (36).

Results

Protein Design and Structure Determination

Understanding the substrate selectivity of CYP2E1 as it plays roles in drug metabolism and both normal and pathophysiology has been hampered by the absence of experimental structural information about this enzyme. In order to facilitate structure determination of this integral membrane protein, CYP2E1 was engineered to remove the single N-terminal transmembrane helix (residues 3–23) and add a C-terminal His_4 tag, as has been recently used to crystallize other mammalian P450 enzymes. The resulting truncated protein is a monotypic membrane protein that can be solubilized, purified, and maintained in an active, monodisperse state in the presence of detergent. Absorbance spectra revealed that the purified CYP2E1 was in a mixed spin state, suggesting that in solution some CYP2E1 protein molecules have

water bound to the sixth coordination site of the heme iron, but in other molecules the water is absent and the heme iron is five-coordinate. Most mammalian P450s are water-coordinated in the resting state, but this mixed spin state has previously been reported for full-length human CYP2E1, although it has also been isolated as either all high spin or all low spin (37-40). When reconstituted with NADPH-cytochrome P450 reductase and cytochrome b_5, the truncated, purified CYP2E1 is catalytically competent in performing two classic CYP2E1 reactions: 2-hydroxylation of *p*-nitrophenol and 6-hydroxylation of chlorzoxazone (data not shown). Additionally, this enzyme has a K_d value for indazole similar to those previously reported for the full-length rabbit CYP2E1 (41). Diffraction data were collected to 2.2 Å on a single crystal of CYP2E1 co-crystallized with indazole. Data collection and refinement statistics are described in Table 4.1. The final CYP2E1/indazole model includes residues K31-S493, with the exception of 138-139. The two unmodeled residues are part of a GXG motif in the loop connecting helices C and D and whose flexibility likely contributed to disorder. Coordinates have been deposited in the Protein Data Bank (PDB code 3E6I).

Table 4.1: Data collection and refinement statistics.

	Indazole-bound
Data collection	
Space group	$P4_3$
Cell dimensions	
a, *b*, *c* (Å)	71.2, 71.2, 225.8
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)*	112.51 - 2.20 (2.26 - 2.20)
R_{sym}*[+]	0.080 (0.338)
I/σI*	15.5 (3.5)
Completeness (%)*	99.7 (99.3)
Redundancy*	5.9 (3.5)
Refinement	
Resolution (Å)	41.78 - 2.20
No. of reflections	53,342
R_{work} / R_{free}	22.0/28.3
No. of atoms	
Protein	7,560
Ligand	18
Heme	86
Water	171
B-factors	
Protein	39.9
Ligand	30.0
Heme	31.2
Water	35.0
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.62

*Values in parentheses are for highest-resolution shell.

[+] $R_{sym} = \sum_{hkl} \sum_i \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum_i I_i(hkl)$

Overall Structure

The CYP2E1 structure exhibits the canonical P450 fold consisting of 12 major α-helices and 4 β-sheets, designated A–L and β1–β4, respectively. Short intervening helices typically found in mammalian membrane P450s are also present and are designated by primes or double primes (Figure 4.1 A). The two CYP2E1 molecules present in the asymmetric unit are nearly identical to each other with a root mean squared deviation (RMSD) for the Cα atoms of 0.22 Å. When compared to other mammalian cytochrome P450 structures in the Protein Data Bank, CYP2E1 has the highest structural similarity to human CYP2A13 (PDB 2P85; RMSD 1.00 Å) (Figures 4.1 B and 4.2) and human CYP2A6 (RMSD 1.03 – 1.05 Å depending on the ligand bound to the CYP2A6 structure).

Active site

In the CYP2E1 complex with indazole, there is clear electron density for the inhibitor in the active site (Figure 4.3). One nitrogen of indazole forms a coordinate covalent bond to the heme iron (N2-Fe, 2.19 Å). The adjacent indazole nitrogen hydrogen bonds with the side chain hydroxyl of T303. Coordination of indazole to the heme iron is consistent with the experimentally observed type II spectral changes that suggest nitrogen replaces water in the sixth coordination position of the iron upon titration with IND (data not shown).

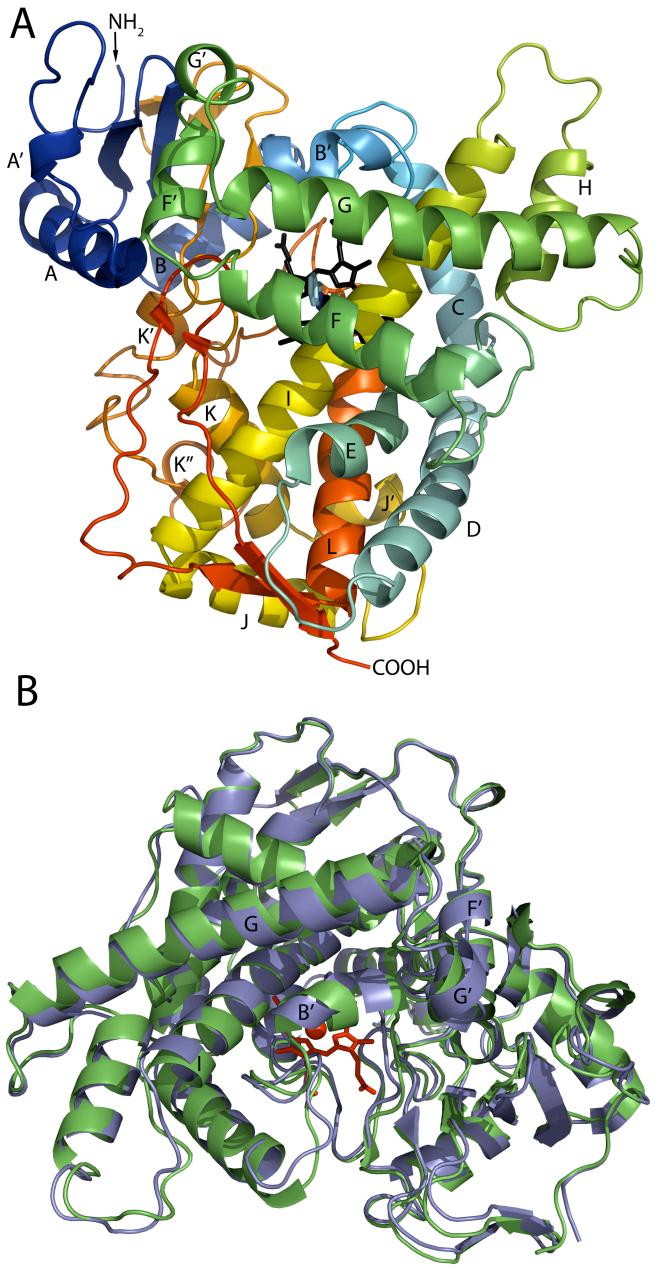


Figure 4.1: Overall structure of CYP2E1 and overlay with CYP2A13. (A) Distal face of CYP2E1 rainbow colored from N-terminus (blue) to C-terminus (red). Helices are labeled using typical P450 nomenclature. (B) Structural overlay of CYP2E1 (blue) and CYP2A13 (green) using secondary-structure matching (42) in COOT (30) with a Cα root mean square deviation of 1.00 Å. All protein structure figures generated using PyMOL (43).

Figure 4.2: Sequences and secondary structures of CYP2E1 and CYP2A13 (2P85) as determined by DSSP (36). Boxes and double underline indicate α helices and β sheets, respectively. Italics indicate N- and C-terminal modifications of CYP2E1 and CYP2A13 for expression, purification, and crystallization.

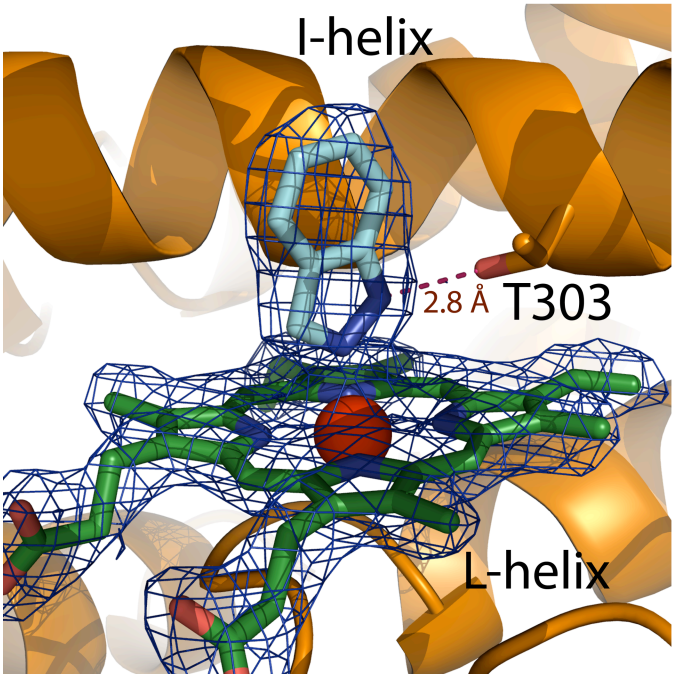


Figure 4.3: Heme and ligand electron density maps. Electron density shown as composite omit σ_A-weighted $2|F_O| - |F_C|$ map contoured at 1.0 σ around the heme and indazole.

In complex with indazole the overall active site cavity of CYP2E1 is small, globular, and highly hydrophobic (Figure 4.4 A). The active site volume was 190 $Å^3$, consistent with the small molecular weight of most CYP2E1 substrates. The amino acid residues that line the active site cavity include A299 and T303 on successive turns of the I helix and which bracket the active site. The opposing face of the active site is formed by the intersection of residues in and near the B' helix (F106, I115, and F116), the loop between K and $\beta_{1\text{-}4}$ (V364 and L368) and the $\beta_{4\text{-}1}/\beta_{4\text{-}2}$ turn (F478) (Figure 4.4 A). A total of five phenylalanine residues (F106, F116, F207, F298, and F478) form the face opposite the heme, many separating the active site from two other enclosed voids.

Second Void Adjacent to Active Site

A second enclosed cavity is found adjacent to the active site cavity (Figure 4.4). This second isolated cavity has a volume of 77 $Å^3$, is located between the I, F, G, and B' helices, and is not present in related P450 enzymes. In CYP2A13 and CYP1A2 a small portion of this area is part of the active site, while protein atoms occupy this area in CYP2A6. The active site and this adjacent cavity are separated by the approach of the side chains of F106 and F298. The narrowest constriction between the two cavities is a distance of 3.8 Å between the atom centers in these two aromatic side chains (Figure 4.4 B), leaving a gap of 0.4 Å between the van der Waals surfaces. Residues lining this additional cavity are H109, L202, F203, N206, V239, V242, K243, A294, and D295.

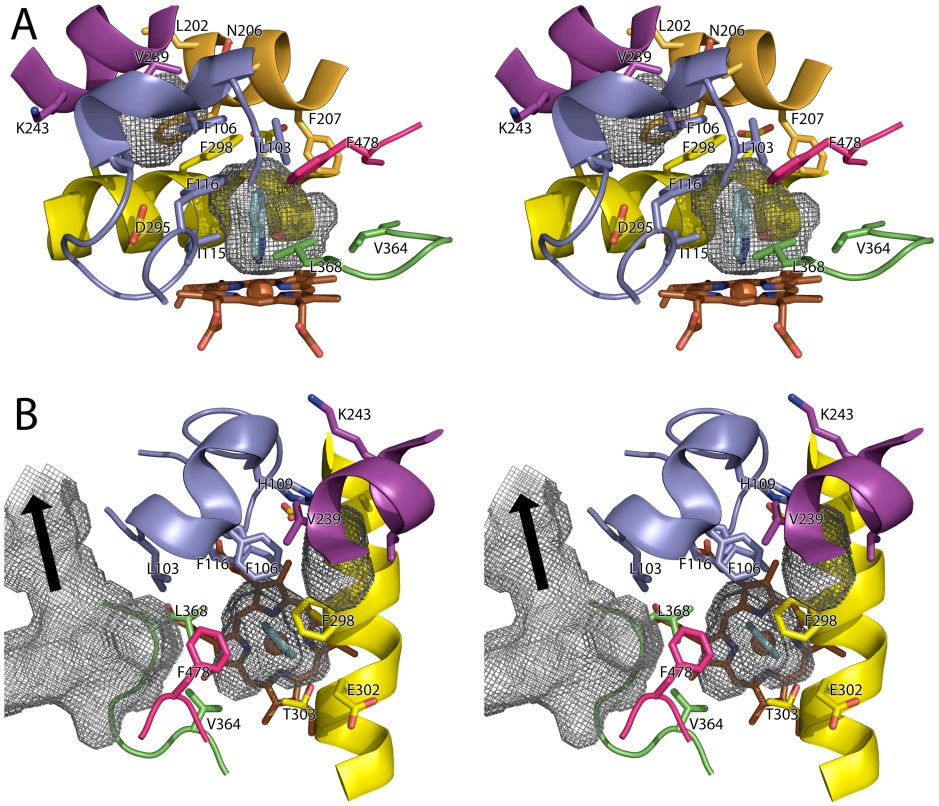


Figure 4.4: Active site of CYP2E1. Wall-eyed stereo views of the CYP2E1 active site illustrating constrictions between the ligand-containing active site, the small distal void, and the substrate access channel. Cavities are shown in grey mesh. Substrate access channel is omitted from panel (A) and F-helix omitted from panel (B) for clarity. Helices and loops colored as indicated: B' helix and adjacent loop (blue); F helix (orange); G helix (purple); I helix (yellow); loop between K and $\beta_{1\text{-}4}$ (green); $\beta_{4\text{-}1}/\beta_{4\text{-}2}$ turn (pink).

Access Channel

CYP2E1 also has a broad channel originating adjacent to the active site opposite the I helix (Figure 4.4 B), and ultimately extending to the surface of the protein. The access channel contains a number of well-defined water molecules. The channel is not linear, but instead has several side channels toward either side of the F helix, though these do not reach the surface. On the protein surface, the single channel opening is surrounded by residues in the loop prior to the B' helix, the loop between the F' and G' helices, and residues from the β1 system (Figure 4.4 A). At its interior terminus, the access channel is separated from the active site cavity primarily by F478 with flanking residues including L103 and active site residues V364 and L368 (Figure 4.4 B). The narrowest constriction is between F478 at the tip of the β4 system and L103 that immediately precedes the B' helix (3.7 Å between atom centers leaving a gap of 0.6 Å between van der Waals surfaces). Slight rotation of F478 would easily allow direct communication between the active site and this channel to allow ligand or solvent access and egress. There is space to either side of F478 for such a rotation in the access channel (Figure 4.5 B).

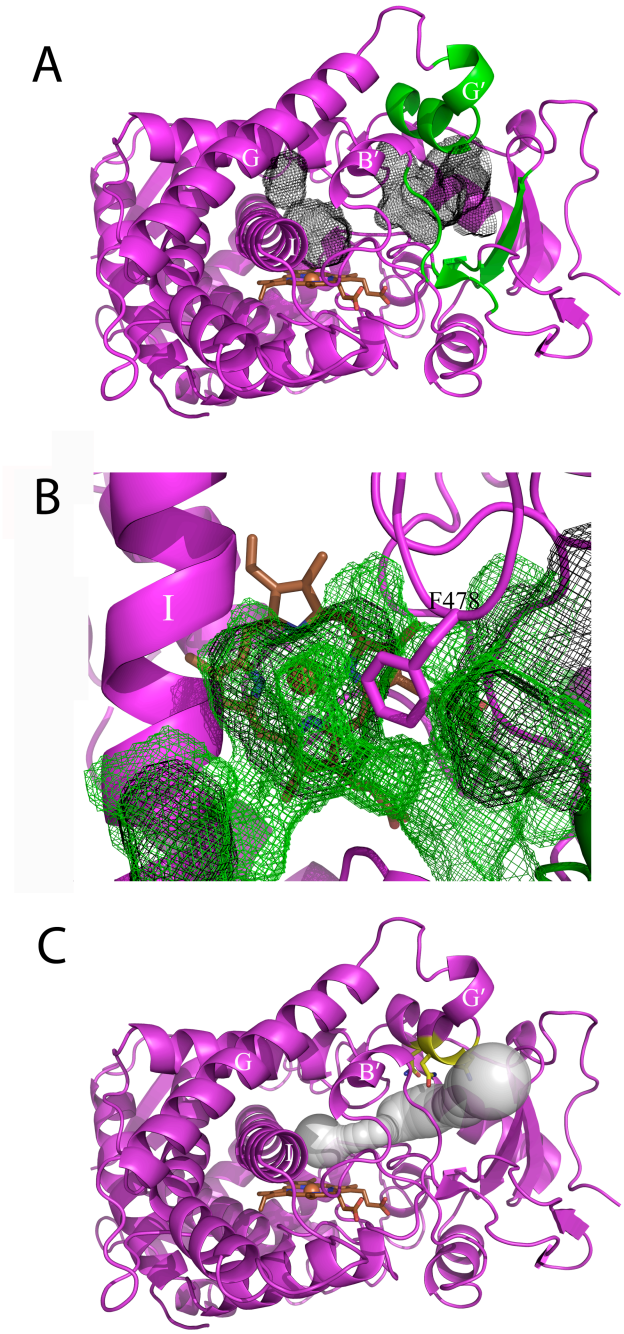


Figure 4.5: CYP2E1 access channel location (A) Access channel, active site, and extra volume are shown in black mesh. Protein regions bordering the access channel entrance are colored in green. (B) Active site void connectivity using a 1.4 Å radius probe (black mesh) versus a 0.9 Å radius probe (green mesh). (C) Most accessible exit path calculated by CAVER (35) is shown as light grey spheres. Residues proposed to interact with a fatty acid carboxylate are shown in yellow.

Discussion

Crystallization

The crystallization of CYP2E1 had been pursued for several years using a variety of ligands but no crystals had been obtained. All of these attempts utilized a buffer system that had been used in the crystallization of all previously published mammalian CYPs. This buffer system was high ionic strength (50 mM KPi, 500 mM NaCl) and included glycerol (20%) as a stabilizing agent. The two major breakthroughs in the CYP2E1 crystallization were the use of indazole as a high affinity ligand and altering the protein buffer system to low ionic strength (100 mM KPi) and exchanging the stabilizing agent for sucrose (0.5 M). This ligand and buffer system appears to help stabilize CYP2E1 as observed in facilitating crystal formation and in helping maintain the characteristic P450 spectrum that can be alter when the protein is not stabilized. This approach may be useful in crystallizing other difficult to crystallize CYPs.

Active Site

The size of the immediate active site (190 $Å^3$) is the smallest observed for a human cytochrome P450 enzyme (Figure 4.6 A) and is consistent with the low molecular weight of many CYP2E1 substrates. CYP2E1-metabolized drugs like the analgesic acetaminophen (MW 151), the muscle relaxant chlorzoxazone (MW 170), the volatile anesthetic sevoflurane (MW 200), and ethanol (MW 46) could all be easily accommodated within this volume, as well as the marker substrate 4-nitrophenol (MW 140). In fact, CYP2E1 has been called a "molecular sieve" (44)

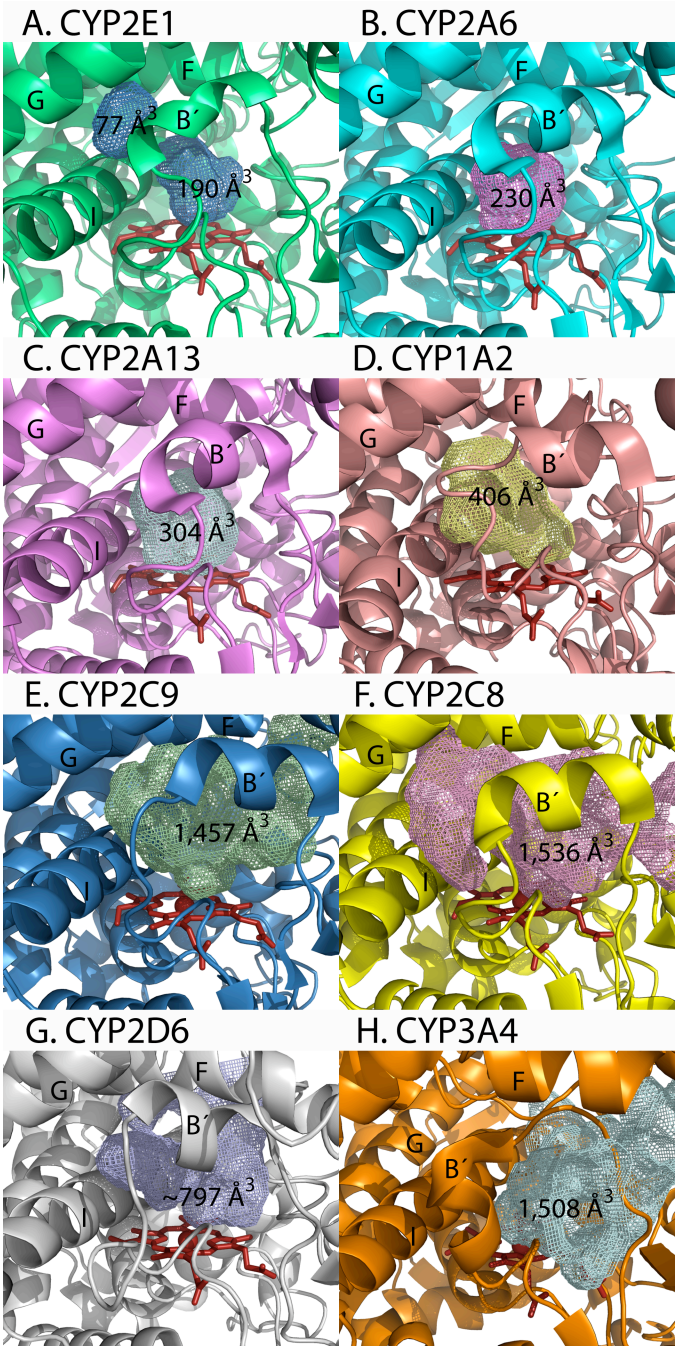


Figure 4.6: Active site and access channel comparisons for human xenobiotic-metabolizing cytochrome P450 enzymes. (A) CYP2E1, (B) CYP2A6 (PDB 1Z10), (C) CYP2A13 (PDB 2P85), (D) CYP1A2 (PDB 2HI4), (E) CYP2C9 (PDB 1OG2), (F) CYP2C8 (PDB 1PQ2), (G) CYP2D6 (PDB 2F9Q), and (H) CYP3A4 (PDB 1TQN). The volumes were calculated by VOIDOO as described in Methods (Chapter 2).

with respect to its preference for small substrates. Although other isozymes may act on these same substrates with a higher V_{max}, CYP2E1 is frequently the low K_m or high affinity enzyme.

Cytochromes P450 2A6, 2A13, and 1A2 also have relatively small active site cavities and typical substrates. The volumes of the CYP2A13 and CYP2A6 cavities are 304 $Å^3$ and 230 $Å^3$, respectively, while the CYP1A2 active site is 406 $Å^3$ (Figure 4.6 B-D). Comparison of the CYP2E1 active site cavity with those of CYP2A6 and CYP2A13 indicates that one of the reasons for the decrease in the active site volume is due to the identity of the residue at position 115, which is directed into the active site. CYP2E1 has Ile at this position, while 2A6 and 2A13 have Val and Ala, respectively. Several residues that line the cavity of CYP2A13 are not part of the CYP2E1 active site cavity because other active site residues screen them. For example, in CYP2E1 H109 and A294 are cut off from the active site cavity by F106 and F116; D295 is obscured from the active site by I115; and P363 is positioned behind V364.

The CYP2E1 active site cavity is most open above the heme D pyrrole ring. This is in general agreement with heme modification studies in which the aryl group of aryl-heme iron complexes predominantly modifies the D ring of CYP2E1 (45,46). In addition, the active site is largely nonpolar with a number of bordering phenylalanine residues that might interact with aromatic substrates. This is consistent with the characterization of many CYP2E1 substrates as neutral compounds with relatively high logP values, many often containing an aromatic ring.

The only polar side chain lining the active site is the highly conserved T303. This residue orients the indazole in the CYP2E1 active site through hydrogen bonding to the side chain hydroxyl. The mutation of T303 in rabbit CYP2E1 to serine, valine, or alanine has shown varying effects on CYP2E1 activity. The mutation T303V is less active in fatty acid hydroxylation than wild type CYP2E1 but retains the fatty acid regioselectivity (47). CYP2E1 T303S also has decreased activity in fatty acid hydroxylation, but additionally loses regioselectivity, suggesting the methyl group is important in orienting the fatty acids in the active site (47,48). Similarly, the CYP2E1 T303A mutation alters the metabolism of isothiocyanates (49). This residue may play a similar role in the positioning of smaller substrates. Evidence suggests that this conserved T303 plays a role in proton delivery to the active site in other P450 enzymes (50-52) and in CYP2E1 (53), but unfortunately no direct studies on the uncoupling of the CYP2E1 T303 mutants have been published.

Several residues in the active site of CYP2E1 have been mutated to the corresponding CYP2B6 residues (54). The CYP2E1 mutants V364L, L368V, and F478V had increased 7-ethoxy-4-trifluoromethylcoumarin (7-EFC) deethylation activity, but lower *p*-nitrophenol (PNP) hydroxylation activity than wild type CYP2E1, indicating these residues are important for CYP2E1 specificity. In the CYP2E1 structure, these three residues line the active site and may play direct roles in positioning these substrates by recontouring the active site volume. These substitutions made in CYP2E1 are likely to change the active site steric bulk to be less compatible with orientation of *p*-nitrophenol for hydroxylation of the phenol ring

than *O*-dealkylation of ethoxycoumarin. In CYP2B6, the combination of substitution of a larger residue at 364 and a smaller residue at 368 may complement each other so that this part of the active site void volume is shifted toward residue 368 (Figure 4.4 B). However, V364, L368, and F478 also form part of the constriction between the active site cavity and the access channel and therefore could also play roles in substrate access. Substitution of F478 to valine would likely eliminate the constriction between the active site and the access channel (Figure 4.4 B). The CYP2E1 mutant L210I also showed less PNP hydroxylation activity than wild type but similar 7-EFC deethylation activities. The residue L210 is not directly part of the active site cavity wall, but rather packs against the aromatic ring of F478 between the access channel and the active site. Thus, altering the amino acid at this position to an isoleucine could have indirect effects on substrate access or on ligand positioning in the active site.

Access Channel

The large channel observed in CYP2E1 nearly connects the protein exterior to the active site void, barring only a constriction formed primarily by the side chain of F478. The narrowest distance between protein carbon atom centers (L103 and F478) is 3.7 Å. There is space on both sides of the F478 side chain into which this aromatic ring could rotate to open up a connection between the two spaces. Even without moving F478, using a smaller spherical probe to map the cavities in VOIDOO shows connections both between the active site and the access channel on either side of F478

and between the active site and the second enclosed void above the I helix (Figure 4.5 B).

The program CAVER was used to find the most accessible path from the active site to the surface of the protein. The most accessible exit/entrance path leaves the active site between L103 and F478 and then follows the VOIDOO access channel to the surface of the protein (Figure 4.5 C). When searching for multiple routes out of the protein, the top four such paths all described the same basic route, transiting from the active site to the access channel on one side or the other of F478.

The immediate 190 $Å^3$ active site is not large enough to accommodate the C_{10}-C_{20} fatty acids that CYP2E1 is known to metabolize. Saturated C12-C18 fatty acids are hydroxylated only at the ω-1 position, with the highest turnover observed for lauric acid (12:0) (55). Additionally, if the carboxylate of palmitic acid (16:0) is methylated, no metabolism occurs (55). These results suggest that when the penultimate alkyl carbon is positioned in the active site for hydroxylation, the carboxylate forms critical interactions with residues 11-15 carbon bond lengths away. The CYP2E1 primary active site is obviously too small to contain the fatty acid, but if the side chain of F478 was rotated slightly, the fatty acid hydrocarbon chain could extend into the substrate access channel where the side chain nitrogens of Q216 and N219 are at the appropriate distance to hydrogen bond with the carboxylate (Figures 4.5 C and 4.7). The side chain of N220 lines the access channel slightly farther away from the active site and might facilitate binding of longer chain fatty acids. All three of these proposed carboxylate-binding residues are completely conserved in CYP2E1

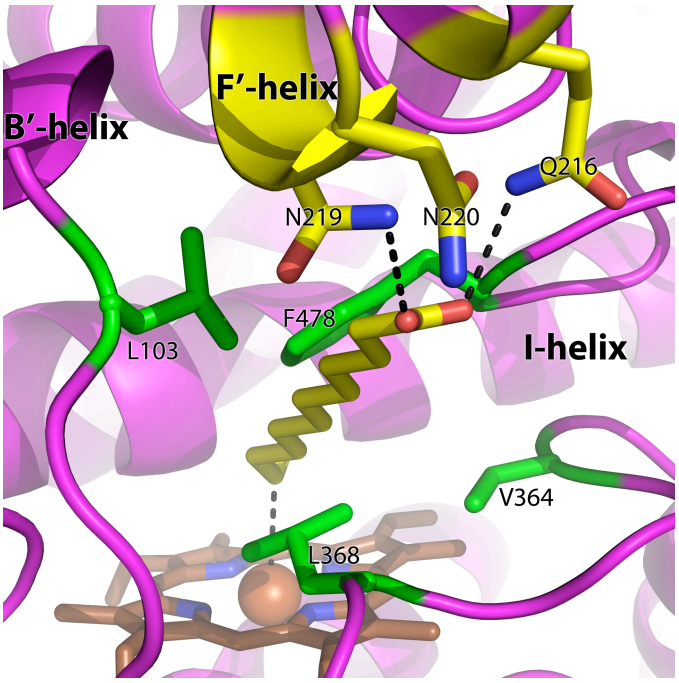


Figure 4.7: Lauric acid structure overlaid on CYP2E1, demonstrating the distance from the active site to the QxxNN access channel residues proposed to bind the carboxylate of fatty acid substrates. In this example the manually docked lauric acid is interacting with Q216 and N219 and the subterminal carbon is in position for the experimentally observed ω-1 hydroxylation.

proteins from different species, forming a QxxNN motif. Additionally, at the level of these polar putative COOH binding residues there is also a large hydrophobic side pocket in the access channel, which might also accommodate parts of the longer chain fatty acids in a more compact, rather than extended, conformation. This proposed orientation is similar to the orientation of fatty acid analogs bound in the structure of BM3, a P450 from *Bacillus megaterium* (56).

CYP2E1 metabolizes the unsaturated fatty acids arachidonic acid and linoleic acid at the terminal or subterminal positions (23), but also forms epoxides from arachidonic acid (14,15-, 11,12- and 8,9- EET (23)) and linoleic acid (9,10- and 12,13- (57)). In the case of arachidonic acid omega-1 hydroxylation, the carboxylate could even bind to Q75 at the opening of the channel, which would still allow extension of the ω-1 carbon over the heme iron (Figure 4.8). However, in order to position the central carbons of arachidonate for the observed epoxide metabolites, the fatty acid would have to penetrate farther into the active site. If the hydrophobic F298 and F106 side chains that form the constriction between the main active site and the "extra" void were also slightly repositioned, the terminus of the alkyl chain could fit into this "extra" space, while the mid-region of the fatty acid could be positioned over the iron for epoxide formation, and the carboxyl group could interact with the putative carboxylate-binding QxxNN residues in the access channel (Figure 4.8). This would require a significant bend in the fatty acid chain, which would be consistent with the structure of unsaturated cis fatty acids. Flash photolysis has shown that arachidonic acid can abolish CO dissociation from the heme, suggesting

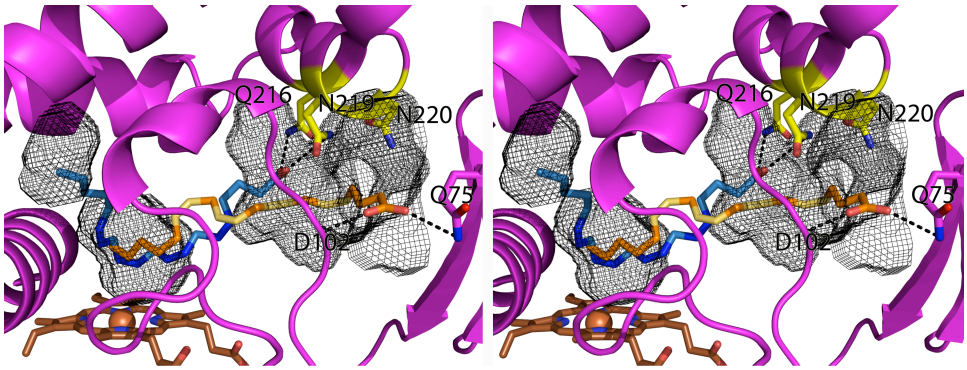


Figure 4.8: Arachidonic acid structures overlaid with CYP2E1. Arachidonic acid manually docked into CYP2E1 in conformations consistent with ω-1 hydroxylation (blue carbons) or 11,12-epoxidation (tan/orange carbons). In this figure the carboxylate was positioned to interact with Q216 and N219 (for epoxidation) or with Q75 and backbone nitrogen of D102 (for omega-1 hydroxylation). The tan and dark blue colored atoms of arachidonic acid indicate the unsaturated positions in the two different binding modes.

that arachidonic acid binding may "plug" up CO exit (58). This physical interpretation of fatty acid binding is consistent with both the observed fatty acid metabolites and the CYP2E1 voids, but should be further investigated by determining the structures of CYP2E1/fatty acid complexes.

Though the human CYP2A enzymes 2A13 and 2A6 are most structurally similar to CYP2E1, neither of them have a substrate access channel in the structures determined to date. Compared to the CYP2A enzymes, the secondary structural elements of CYP2E1 are overlaid closely throughout the core of the protein, with seemingly subtle shifts localized in one quadrant of the protein consisting of the F' and G' helices, the B' helix and its adjacent loops, and the N-terminus of the C helix (Figure 4.1 B). However, these structural units partially frame the opening of the access channel, along with residues from the β1 system. The CYP2E1 B' helix is a 3_{10} helix and is shorter than in CYP2A13 by one residue due to the deletion of a tryptophan that is present in CYP2A13 sequence (W109). Both the B' and G helices are slightly shifted toward each other, and in combination with small shifts in the F' and G' helices, create the CYP2E1 access channel. However, a variety of access channels have been shown in other mammalian cytochromes and obviously the substrates must penetrate to the active site and metabolites must diffuse out (Figure 4.5). CYP2C8 has two access channels, one on each side of the B' helix (59). The channel in CYP2C9 exits between helices F and I and the C-terminal β_4 sheet system (37). CYP2D6 has the same hydrophilic access channel as CYP2C9 and a second putative channel exiting between the G and I helices (60). CYP3A4 contains three

access channels, a channel in the same area as CYP2E1, a second channel through the middle of the B' loop, and a third in the same position as CYP2C9 (61,62).

Location of Naturally Occurring Polymorphisms

By comparison with other cytochromes P450, CYP2E1 is fairly highly conserved across species, which has been attributed to its physiological roles. In humans, only three nonsynonymous polymorphisms have been reported for human CYP2E1: R76H, V179I, and V389I (http://www.cypalleles.ki.se). Mutations at positions V179 and V389 do not impact enzymatic activity (63,64) and are buried in the CYP2E1 structure indicating no direct involvement in catalysis. However, R76H showed similar mRNA levels to the wild type protein, but only ~30% of the protein levels (by western blot analysis), suggesting decreased translation or protein stabilization. R76 is part of the $\beta_{1\text{-}2}$ structural unit, but is located on the surface with the side chain oriented toward solvent.

Importance of B´ Helix

Residues flanking and in the B' helix play key roles in forming the opening of the substrate access channel, forming part of the constriction between the access channel and the active site, serving to define the active site, forming part of the constriction between the active site and the second enclosed volume above the active site, and also forming part of the wall of this latter void. The B' region of the protein has few interactions with the remainder of the protein; most interactions are within the B' helix unit. Significant interactions between the B' helix and the remainder of the protein involve only H109 and R112. The H109 imidazole nitrogen hydrogen

bonds to the side chain of D295 in the I helix. Additionally, in molecule A a water molecule resides in this second pocket and is the intermediate link in a hydrogen bonding network between the side chain of H109 in the B' helix on one side of the void and both the carbonyl of V239 and the backbone nitrogen of K243 in the G helix on the opposite side of the void. In a second interaction R112 hydrogen bonds with the side chain of D287 in the I helix. In other cytochromes P450 the B' helix adopts a number of different conformations and may exhibit flexibility in response to various ligands. Thus, residues in this region and their interactions may be critical for protein stability and encapsulating the active site.

Interactions with Electron Delivery Partners NADPH-cytochrome P450 Reductase and Cytochrome b_5

Analysis of the electrostatic surface of CYP2E1 shows a prominent positively charged "bowl" on the proximal side of the enzyme (Figure 4.9). This region is proposed to interact with negatively charged regions of both redox partners: NADPH-cytochrome P450 reductase and cytochrome b_5. Multiple roles have been proposed for cytochrome b_5 in P450-mediated metabolism, including delivering to P450 the second electron required for catalysis and/or stabilization of P450 in support of catalysis. Cytochrome b_5 addition strongly stimulates CYP2E1 metabolism, ~12-fold for *p*-nitrophenol (65), 25-fold for acetaminophen (66), 67-fold for 7-ethoxycoumarin (67), and 270-fold for aniline (68). Although for many P450 proteins stimulation of metabolism can be observed upon addition of either cytochrome b_5 or apo cytochrome b_5 (lacking the heme and therefore unable to do

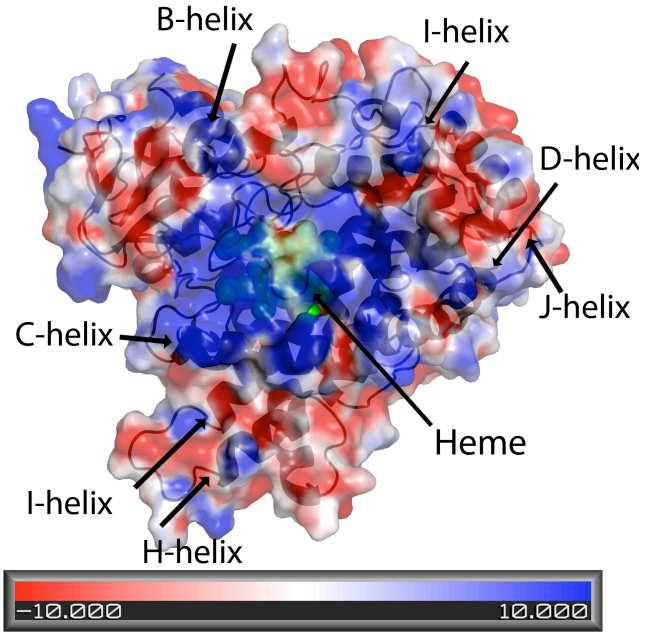


Figure 4.9: Electrostatic surface of CYP2E1. Electrostatic surface of CYP2E1 as calculated by APBS (70) showing the positively charged surface proposed to be the cytochrome b_5/NADPH-cytochome P450 reductase binding site. Arrows marking the location of helices are oriented along the long axes of helices to aid in spatial orientation. Heme shown as green spheres.

electron transfer), apo cytochrome b_5 cannot replace holo cytochrome b_5 in CYP2E1 catalysis, suggesting a role in electron transfer (67). The putative protein interface between CYP2E1 and cytochrome b_5 has been explored by cross-linking and site-directed mutagenesis (69). These studies identified two cross-links: between K428 in CYP2E1 and D53 in cytochrome b_5 and between K434 in CYP2E1 and E56 in cytochrome b_5. The proposed interaction surface consists of the CYP2E1 meander, β-bulge, C, L, and J' helices on the proximal side of the protein and that binds cytochrome b_5 with its heme perpendicular to the CYP2E1 heme. However, the K428 and K434 residues are 14 Å apart on the CYP2E1 structure while the proposed corresponding residues on cytochrome b_5 (D53 and E56) are only ~3.4 Å apart (PDB 1JEX). Of the nine total intermolecular electrostatic interactions proposed by docking (69), only one (R344) is buried in the CYP2E1 structure, while the remainder are surface exposed and do contribute to the positively charged bowl on the proximal surface.

Uncoupling and the production of Reactive Oxygen Species

Very little is known of the structural basis for the propensity of CYP2E1 to produce ROS. In general, reactive oxygen species are generated during P450 catalysis when either the ferrous oxy species decays to produce superoxide or the hydroperoxy form is protonated to release hydrogen peroxide. Collapse of these species could occur if electron or proton delivery is delayed or if the substrate is not positioned for attack. It has been proposed that during uncoupling, the substrate could migrate away from the heme and the activated oxygen thus allowing the

oxygen to react in other ways (68). In CYP2E1, transient connections between the active site and the other voids in the protein might facilitate substrate movement away from the heme after initial binding. Similarly, the presence of water molecules in the CYP2E1 access channel, and the connections between the active site and the access channel may also allow water ready access to the active site. In the resting enzyme, facile movement of solvent into and out of the active site may result in the mixed spin state observed for CYP2E1. This increase in high spin character, which in other P450s has shown a raising of redox potential (70), may then facilitate the generation of ROS. Finally, the particularly unstable nature of the CYP2E1 enzyme may also contribute to uncoupled substrate binding and catalysis. The magnitude and implications of CYP2E1-generated ROS in human disease remain to be elucidated.

Conclusions

In summary, this structure has suggested the structural properties that allow for the metabolism of the seemingly divergent substrate classes metabolized by CYP2E1: small molecular weight compounds sequestered within a small active site and endogenous fatty acids binding with the carboxylate end extending into the access channel. Additional CYP2E1 structures in complex with substrates or fatty acids should be pursued to validate and determine the details of these binding modes. Finally, further studies will be required to understand the propensity of CYP2E1 to generate reactive oxygen species.

References

1. Ioannides, C., and Lewis, D. F. (2004) *Curr. Top. Med. Chem.* **4**(16), 1767-1788
2. Bieche, I., Narjoz, C., Asselah, T., Vacher, S., Marcellin, P., Lidereau, R., Beaune, P., and de Waziers, I. (2007) *Pharmacogenet. Genomics* **17**(9), 731-742
3. Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., and Guengerich, F. P. (1994) *J. Pharmacol. Exp. Ther.* **270**(1), 414-423
4. Ding, X., and Kaminsky, L. S. (2003) *Annu. Rev. Pharmacol. Toxicol.* **43**, 149-173
5. Nanji, A. A., Zhao, S., Sadrzadeh, S. M., Dannenberg, A. J., Tahan, S. R., and Waxman, D. J. (1994) *Alcohol Clin. Exp. Res.* **18**(5), 1280-1285
6. Barnett, C. R., Rudd, S., Flatt, P. R., and Ioannides, C. (1993) *Biochemical pharmacology* **45**(2), 313-319
7. Raucy, J. L., Lasker, J. M., Kraner, J. C., Salazar, D. E., Lieber, C. S., and Corcoran, G. B. (1991) *Mol. Pharmacol.* **39**(3), 275-280
8. Johansson, I., Ekstrom, G., Scholte, B., Puzycki, D., Jornvall, H., and Ingelman-Sundberg, M. (1988) *Biochemistry* **27**(6), 1925-1934
9. Bell, L. C., and Guengerich, F. P. (1997) *J. Biol. Chem.* **272**(47), 29643-29651
10. Lu, Y., and Cederbaum, A. I. (2008) *Free Radical Biol. Med.* **44**(5), 723-738
11. Dostalek, M., Hardy, K. D., Milne, G. L., Morrow, J. D., Chen, C., Gonzalez, F. J., Gu, J., Ding, X., Johnson, D. A., Johnson, J. A., Martin, M. V., and Guengerich, F. P. (2008) *J. Biol. Chem.* **283**(25), 17147-17157
12. Patten, C. J., Thomas, P. E., Guy, R. L., Lee, M., Gonzalez, F. J., Guengerich, F. P., and Yang, C. S. (1993) *Chemical research in toxicology* **6**(4), 511-518
13. Raucy, J. L., Lasker, J. M., Lieber, C. S., and Black, M. (1989) *Archives of biochemistry and biophysics* **271**(2), 270-283
14. Kharasch, E. D., Hankins, D., Mautz, D., and Thummel, K. E. (1996) *Lancet* **347**(9012), 1367-1371
15. Larson, A. M., Polson, J., Fontana, R. J., Davern, T. J., Lalani, E., Hynan, L. S., Reisch, J. S., Schiodt, F. V., Ostapowicz, G., Shakil, A. O., and Lee, W. M. (2005) *Hepatology* **42**(6), 1364-1372
16. Bronstein, A. C., Spyker, D. A., Cantilena, L. R., Jr., Green, J., Rumack, B. H., and Heard, S. E. (2007) *Clin. Toxicol. (Phila)* **45**(8), 815-917
17. Ray, D. C., and Drummond, G. B. (1991) *Br. J. Anaesth.* **67**(1), 84-99
18. Guengerich, F. P., Kim, D. H., and Iwasaki, M. (1991) *Chemical research in toxicology* **4**(2), 168-179
19. Koop, D. R., and Coon, M. J. (1986) *Alcohol Clin. Exp. Res.* **10**(6 Suppl), 44S-49S
20. Koop, D. R. (1986) *Mol. Pharmacol.* **29**(4), 399-404

21. Tassaneeyakul, W., Birkett, D. J., McManus, M. E., Tassaneeyakul, W., Veronese, M. E., Andersson, T., Tukey, R. H., and Miners, J. O. (1994) *Biochemical pharmacology* **47**(10), 1767-1776
22. Peter, R., Bocker, R., Beaune, P. H., Iwasaki, M., Guengerich, F. P., and Yang, C. S. (1990) *Chemical research in toxicology* **3**(6), 566-573
23. Laethem, R. M., Balazy, M., Falck, J. R., Laethem, C. L., and Koop, D. R. (1993) *J. Biol. Chem.* **268**(17), 12912-12918
24. Roy, U., Joshua, R., Stark, R. L., and Balazy, M. (2005) *Biochem. J.* **390**(Pt 3), 719-727
25. Salazar, D. E., Sorge, C. L., and Corcoran, G. B. (1988) *Biochemical and biophysical research communications* **157**(1), 315-320
26. Song, B. J., Matsunaga, T., Hardwick, J. P., Park, S. S., Veech, R. L., Yang, C. S., Gelboin, H. V., and Gonzalez, F. J. (1987) *Mol. Endocrinol.* **1**(8), 542-547
27. Leslie, A. G. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62**(Pt 1), 48-57
28. Evans, P. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62**(Pt 1), 72-82
29. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40**(4), 658-674
30. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2126-2132
31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**(Pt 3), 240-255
32. Vriend, G. (1990) *J. Mol. Graphics* **8**(1), 52-56, 29
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**(2), 283-291
34. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**(Pt 2), 178-185
35. Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006) *BMC Bioinf.* **7**, 316
36. Kabsch, W., and Sander, C. (1983) *Biopolymer* **22**, 2577-2637
37. Wrighton, S. A., Thomas, P. E., Molowa, D. T., Haniu, M., Shively, J. E., Maines, S. L., Watkins, P. B., Parker, G., Mendez-Picon, G., Levin, W., and et al. (1986) *Biochemistry* **25**(22), 6731-6735
38. Lasker, J. M., Raucy, J., Kubota, S., Bloswick, B. P., Black, M., and Lieber, C. S. (1987) *Biochemical and biophysical research communications* **148**(1), 232-238
39. Gillam, E. M., Guo, Z., and Guengerich, F. P. (1994) *Archives of biochemistry and biophysics* **312**(1), 59-66
40. Patten, C. J., and Koch, P. (1995) *Archives of biochemistry and biophysics* **317**(2), 504-513
41. Collom, S. L., Laddusaw, R. M., Burch, A. M., Kuzmic, P., Perry, M. D., Jr., and Miller, G. P. (2008) *J. Biol. Chem.* **283**(6), 3487-3496
42. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2256-2268

43. DeLano, W. L. (2002) PyMOL Molecular Graphics System at www.pymol.org. In.
44. Guengerich, F. P., and Shimada, T. (1991) *Chemical research in toxicology* **4**(4), 391-407
45. Swanson, B. A., Dutton, D. R., Lunetta, J. M., Yang, C. S., and Ortiz de Montellano, P. R. (1991) *J. Biol. Chem.* **266**(29), 19258-19264
46. Mackman, R., Guo, Z., Guengerich, F. P., and Ortiz de Montellano, P. R. (1996) *Chemical research in toxicology* **9**(1), 223-226
47. Fukuda, T., Imai, Y., Komori, M., Nakamura, M., Kusunose, E., Satouchi, K., and Kusunose, M. (1993) *J. Biochem.* **113**(1), 7-12
48. Fukuda, T., Imai, Y., Komori, M., Nakamura, M., Kusunose, E., Satouchi, K., and Kusunose, M. (1994) *J. Biochem.* **115**(2), 338-344
49. Moreno, R. L., Goosen, T., Kent, U. M., Chung, F. L., and Hollenberg, P. F. (2001) *Archives of biochemistry and biophysics* **391**(1), 99-110
50. Imai, M., Shimada, H., Watanabe, Y., Matsushima-Hibiya, Y., Makino, R., Koga, H., Horiuchi, T., and Ishimura, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**(20), 7823-7827
51. Vaz, A. D., Pernecky, S. J., Raner, G. M., and Coon, M. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**(10), 4644-4648
52. Vaz, A. D., McGinnity, D. F., and Coon, M. J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**(7), 3555-3560
53. Blobaum, A. L., Kent, U. M., Alworth, W. L., and Hollenberg, P. F. (2004) *J. Pharmacol. Exp. Ther.* **310**(1), 281-290
54. Spatzenegger, M., Liu, H., Wang, Q., Debarber, A., Koop, D. R., and Halpert, J. R. (2003) *J. Pharmacol. Exp. Ther.* **304**(1), 477-487
55. Adas, F., Salaun, J. P., Berthou, F., Picart, D., Simon, B., and Amet, Y. (1999) *J. Lipid Res.* **40**(11), 1990-1997
56. Haines, D. C., Chen, B., Tomchick, D. R., Bondlela, M., Hegde, A., Machius, M., and Peterson, J. A. (2008) *Biochemistry* **47**(12), 3662-3670
57. Moran, J. H., Mitchell, L. A., Bradbury, J. A., Qu, W., Zeldin, D. C., Schnellmann, R. G., and Grant, D. F. (2000) *Toxicol. Appl. Pharmacol.* **168**(3), 268-279
58. Smith, S. V., Koley, A. P., Dai, R., Robinson, R. C., Leong, H., Markowitz, A., and Friedman, F. K. (2000) *Biochemistry* **39**(19), 5731-5737
59. Schoch, G. A., Yano, J. K., Sansen, S., Dansette, P. M., Stout, C. D., and Johnson, E. F. (2008) *J. Biol. Chem.* **283**(25), 17227-17237
60. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006) *J. Biol. Chem.* **281**(11), 7614-7622
61. Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., and Jhoti, H. (2004) *Science (New York, N.Y* **305**(5684), 683-686
62. Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J. Biol. Chem.* **279**(37), 38091-38094

63. Hu, Y., Oscarson, M., Johansson, I., Yue, Q. Y., Dahl, M. L., Tabone, M., Arinco, S., Albano, E., and Ingelman-Sundberg, M. (1997) *Mol. Pharmacol.* **51**(3), 370-376
64. Fairbrother, K. S., Grove, J., de Waziers, I., Steimel, D. T., Day, C. P., Crespi, C. L., and Daly, A. K. (1998) *Pharmacogenetics* **8**(6), 543-552
65. Chen, W., Peter, R. M., McArdle, S., Thummel, K. E., Sigle, R. O., and Nelson, S. D. (1996) *Archives of biochemistry and biophysics* **335**(1), 123-130
66. Chen, W., Koenigs, L. L., Thompson, S. J., Peter, R. M., Rettie, A. E., Trager, W. F., and Nelson, S. D. (1998) *Chemical research in toxicology* **11**(4), 295-301
67. Yamazaki, H., Nakamura, M., Komatsu, T., Ohyama, K., Hatanaka, N., Asahi, S., Shimada, N., Guengerich, F. P., Shimada, T., Nakajima, M., and Yokoi, T. (2002) *Protein Expression Purif.* **24**(3), 329-337
68. Porter, T. D. (2002) *J. Biochem. Mol. Toxicol.* **16**(6), 311-316
69. Gao, Q., Doneanu, C. E., Shaffer, S. A., Adman, E. T., Goodlett, D. R., and Nelson, S. D. (2006) *J. Biol. Chem.* **281**(29), 20404-20417
70. Das, A., Grinkova, Y. V., and Sligar, S. G. (2007) *J. Am. Chem. Soc.* **129**(45), 13778-13779

Chapter 5.

Structure of CYP2E1 Bound with the Fatty Acid Analog, ω-Imidazolyl-decanoic acid

The atomic coordinates and structure factors of the structure described in this chapter have been deposited in the protein data bank (code 3GPH).

Introduction

Cytochrome P450 (CYP) is a superfamily of heme-containing proteins that catalyze the monooxygenation of a wide variety of substrates. One substrate class of CYP is fatty acids. The metabolism of fatty acids by CYP has been proposed to be important in maintaining homeostasis of fatty acid levels. This may be especially important in the metabolism of mid-range length fatty acids liberated from the peroxisomal and mitochondrial β-oxidation pathway (1). The reactions known to be catalyzed by CYP are the ω-hydroxylation or (ω-1)-hydroxylation of saturated fatty acids and epoxidation of the alkenes of unsaturated fatty acids. One CYP known to metabolize fatty acids in addition to the oxidation of xenobiotics is CYP2E1.

There is a great wealth of biochemical information about CYP metabolism and binding of fatty acids but little structural information to date. The structural information currently in hand is for the P450 domain of the fatty acid hydroxylase of *Bacillus megaterium* (P450BM3) bound to either *N*-palmitoylglycine (2), palmitoleate (to a mutant enzyme) (3), or *N*-(10-imidazolyldodecanoyl)-*L*-leucine (4). In the first two cases the structures show protein/fatty acid complexes in conformational states that would not be the catalytic form (the ligands are too far

from the heme). In the case of the imidazole-based compound, the imidazole end of the fatty acid analog is coordinate covalently bound to the heme iron, locking it in place. In mammalian systems there are no crystal structures with fatty acids bound in the active site of the CYP. Biochemical data to date has given insights into distances between the heme and hydrogen bond donors that would presumably interact with the carboxylate end of fatty acids, but the identity of those donors remains elusive.

Heme-binding fatty acid analogs similar to *N*-(10-imidazolyldodecanoyl)-*L*-leucine were originally designed in the lab of Dr. Robert Hanzlik for the inhibition of lauric acid ω-hydroxylation by CYP4A1 (5-8). Many "Lewis base" type functional groups were screened with the most effective heme-binding analogs being ω-imidazolyl fatty acid analogs (8). These compounds were high affinity and gave a characteristic spectral shift, indicative of direct nitrogen binding to the heme iron. Though CYP2E1 metabolizes fatty acids with chain lengths from C9 (9) to C20 (10), the fatty acid metabolized at the highest rate for both CYP4A1 and CYP2E1 is lauric acid. This suggests that the ligands designed to probe the CYP4A1 active site might also bind to CYP2E1.

Recently, the structure of CYP2E1 has been solved and a fatty acid binding site proposed that involves the residues Q216, N219, or N220 acting as hydrogen bond donors to the carboxylate of fatty acids (see Chapter 4, or (11)). However, further work was needed to confirm or disconfirm the identity of the residues involved in fatty acid binding. Initially two approaches were pursued to answer this problem including mutagenesis of the proposed residues and X-ray crystal structures

of CYP2E1 bound with a fatty acid analog. Mutants were generated by site-direct mutagenesis at the proposed locations, mutating the polar amino acids to either leucine or alanine to remove H-bonding or H-bonding and bulk, respectively. Unfortunately, inconsistent results were obtained from mutated purified proteins with regard to spectral characteristics indicating the ligand bound state, as well as the presence of inactive protein (P420). To accomplish the task of understanding fatty acid binding, ω-imidazolyl-fatty acid analogs were generated, binding affinities were determined, and a cocrystal structure of CYP2E1 solved bound to one of these fatty acid analogs, ω-imidazolyl-decanoic acid. This work identified a fatty acid carboxylate binding mode distinct from that proposed based on the previous structure.

Methods

Ligand Synthesis

The synthesis of ω-imidazolyl-hexanoic acid, ω-imidazolyl-octanoic acid, ω-imidazolyl-decanoic acid, and ω-imidazolyl-dodecanoic acid was performed as previously described (5,8) with the following modifications. Methylation of the bromo-fatty acids was performed using thionyl chloride and methanol rather than diazomethane. Purification of ω-imidazolyl-hexanoic acid, ω-imidazolyl-octanoic acid, and ω-imidazolyl-decanoic acid was performed using anion exchange (Dowex 1X8 200, formate counter ion (12)) rather than cation exchange. Finally, purification of ω-imidazolyl-dodecanoic acid was performed by crystallization as the lithium salt.

Spectral Binding Titrations

See Methods (Chapter 2) for the elaboration of spectral binding titrations.

Protein Engineering, Expression, Purification, and Ligand Complexation

See Methods (Chapter 2) for the description of protein engineering, expression, purification, and formation of the ligand complex.

Protein Crystallization, Data Collection, and Structure Determination

Crystals were grown by hanging drop vapor diffusion as described in Methods (Chapter 2). The CYP2E1/ω-imidazolyl-decanoic acid complex (0.91 mM CYP2E1, 25 mM ω-imidazolyl-decanoic acid) was equilibrated against 0.1 M NaHEPES pH 7.5, 5% *iso*-propanol, and 20% PEG 2000 MME. Crystallization drops were initially seeded with microcrystals (13) of the CYP2E1/indazole complex to initiate nucleation of CYP2E1/ω-imidazolyl-decanoic acid crystals. In order to ensure crystal homogeneity and absence of indazole, portions of the resulting crystals were used to serially seed fresh crystallization drops. Crystals were immersed in 0.1 M NaHEPES pH 7.5, 5% *iso*-propanol, 30% PEG 2000 MME as a cryoprotectant before being flash cooled in liquid nitrogen for data collection. A single native data set was collected on beamline 17-BM at the Advanced Photon Source (Argonne, IL) with a wavelength of 1.00 Å at 100 K. Data was processed using DENZO (14) and SCALEPACK (15). The CYP2E1/ω-imidazoyl-decanoic acid structure was solved by molecular replacement using the program PHASER (16) with a search model consisting of CYP2E1/indazole (PDB 3E6I). Iterative model building and refinement were performed using COOT (17) and REFMAC (18).

Structure validation was performed using WHAT-IF (19) and PROCHECK (20). The Ramachandran plot showed 86.5% of the amino acid residues in the most favorable region, 12.1% in the additional allowed region, 0.9% in the generously allowed region, and 0.5% in the disallowed region. The residues in the disallowed region are I41 (molecule A and B), R379 (molecule A only), and Q141 (molecule A only). I41 and R379 were also in the disallowed region in the indazole-bound structure.

Structure Analysis

Voids and exit channels were probed using VOIDOO (21) and CAVER (22) as described in Methods (Chapter 2).

Results

Synthesis and Ligand Binding

ω-Imidazolyl-fatty acid analogs were prepared essentially as described (5,8) with even chain lengths from C6 (ω-imidazolyl-hexanoic acid) to C12 (ω-imidazolyl-dodecanoic acid). Only the simple modifications described in the methods were made to the synthesis, largely to avoid the use of diazomethane (Scheme 5.1). The spectral binding affinities of fatty acid analogs are listed in Table 5.1. Upon binding these compounds cause the heme soret peak to shift from a split peak at 390 nm and 414 nm to a single peak at 424 nm characteristic of a Type II ligand (nitrogen bound directly to the heme iron). The ligand with the highest affinity for CYP2E1 is ω-imidazolyl-dodecanoic acid with a K_D of 2 μM. Spectral binding affinities decrease

Scheme 5.1: Synthesis of ω-imidazolyl-fatty acid analogs

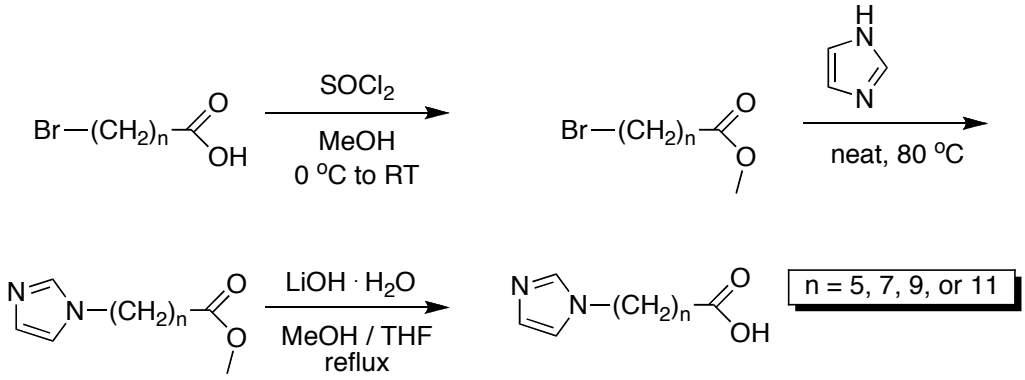
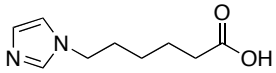
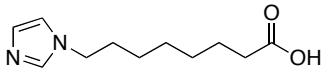
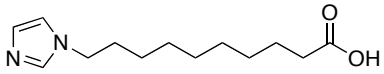
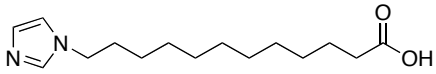


Table 5.1: Ligand binding data. Duplicate lines for each compound corresponds to separate titrations, which are averaged in the last column.

Ligand	ClogP[#]	K_D (μM)	Δ A_{max}	R^{2+}	Average K_D (μM)
	0.69	No Shift*			
		No Shift*			
	1.7	21	0.011	0.97	
		21	0.018	0.97	21
	2.8	8.0	0.016	0.98	
		10.	0.029	0.98	9.0
	3.7	1.8	0.022	0.96	
		1.9	0.021	0.97	1.8

*No shift: No spectral shift observed at ligand concentration up to 230 μM ligand.

[+]R^2: Residual for fit to $\Delta A = \frac{\Delta A_{max} \times [ligand]}{K_D + [ligand]}$.

[#]ClogP: calculated log of the octanol / water partitioning coefficient.

with shorter chain lengths for the analogs examined with ω-imidazolyl-hexanoic acid exhibiting no observable binding up to 230 μM.

Structure Determination and Overall Fold

The cocrystal structure of CYP2E1/ω-imidazolyl-decanoic acid was solved at 2.7 Å using the CYP2E1/indazole (PDB 3E6I) structure as the search model. The structure was refined to give the final statistics shown in Table 5.2. The final model was very similar to that of CYP2E1/indazole with a R.M.S.D. of Cα atoms of 0.35 Å (using secondary structure matching (23)). As observed in the CYP2E1/indazole structure, the Gly-X-Gly residues 138-139 were omitted from the final model due to weak electron density. With this exception the model encompasses residue 31 through 493.

Ligand binding interactions

The electron density ($2|F_O| - |F_C|$ map shown in Figure 5.1) of ω-imidazolyl-decanoic acid bound in the active site of CYP2E1 is continuous from the heme-binding imidazole to the carboxylate. The carboxylate interacts with the side chain of N206 in the active site of both molecules in the asymmetric unit with hydrogen bond distances of 2.5 Å in molecule A and 3.3 Å in molecule B. On the imidazole end of the molecule the non-alkylated imidazole nitrogen is coordinate covalently bound to the heme iron with distances of 2.4 Å in molecule A and 2.3 Å in molecule B. This is consistent with nitrogen-iron distances observed for other CYP enzymes in complex with Type II ligands. The alkylated imidazole nitrogen lies outside of hydrogen bonding distance from the CYP2E1 hydrogen bond

Table 5.2: Data collection and refinement statistics.

	ω-imidazolyl-decanoic acid-bound
Data collection	
Space group	$P4_3$
Cell dimensions	
a, *b*, *c* (Å)	70.7, 70.7, 222.8
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)*	50.00 – 2.70 (2.80-2.70)
R_{merge}*	0.104 (0.457)
% > 3 σ (*I*)*	76.5 (48.5)
Completeness (%)*	95.4 (93.4)
Redundancy*	6.7 (6.3)
Refinement	
Resolution (Å)	31.31 – 2.70
No. reflections	27040
R_{work} / R_{free}	21.2 / 28.3
No. atoms	
Protein	7536
Ligand	34
Heme	86
Water	25
B-factors	
Protein	28.9
Ligand	29.9
Heme	24.8
Water	14.8
R.m.s. deviations	
Bond lengths (Å)	0.019
Bond angles (°)	1.909

*Values in parentheses are for highest-resolution shell

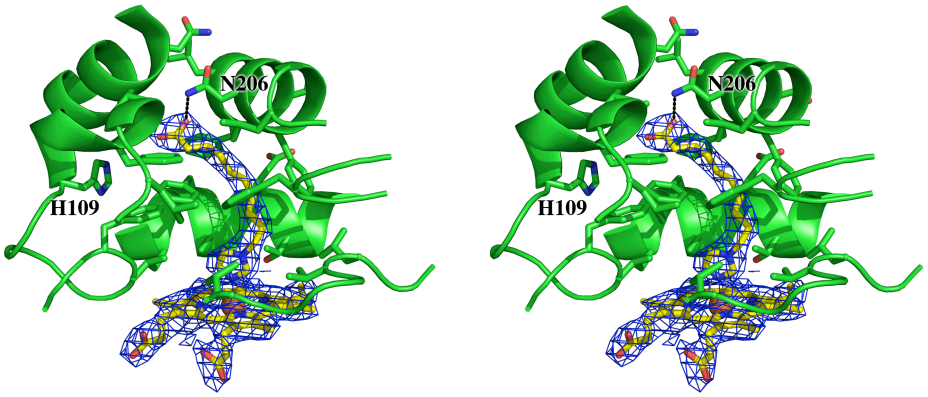


Figure 5.1: Heme and ligand electron density maps shown in wall-eyed stereo view. Electron density shown as composite omit σ_A-weighted $2|F_O| - |F_C|$ map contoured at 1.0 σ around the heme and ω-imidazolyl-decanoic acid.

donor/acceptor, T303. The lack of H-bonding by this imidazole nitrogen is logical since the lone pair electrons of the alkylated imidazole nitrogen would be delocalized in the aromatic ring and thus decrease its affinity for accepting hydrogen bonds.

Active site

The active site of CYP2E1 makes only subtle changes in order to accommodate the fatty acid analog, ω-imidazolyl-decanoic acid. In comparing the CYP2E1 structure bound with indazole or with ω-imidazolyl-decanoic acid (Figure 5.2) the largest change is the position of F298. This side chain forms a barrier between the active site void and small additional void in the indazole-bound structure. In the ω-imidazolyl-decanoic acid-bound structure, this side chain rotates approximately 60°, making the two voids connect and allowing the analog to extend into the space between the B′, and F-helices beneath the G-helix. It interacts with N206, which has not altered positions between structures. Several other residues also make minor shifts to acclimate to ω-imidazolyl-decanoic acid, including F106, F207, and F478. These shifts appear to be either in adjustment to the rotation of F298 or the presence of the hydrophobic methylene chain of the ligand. T303 also shifts from its position in the indazole-bound structure. The shifting of side chains in the fatty acid analog-bound structure caused an increase in active site volume from 190 $Å^3$ (indazole-bound) to 445 $Å^3$ (ω-imidazolyl-decanoic acid-bound) (Figure 5.3).

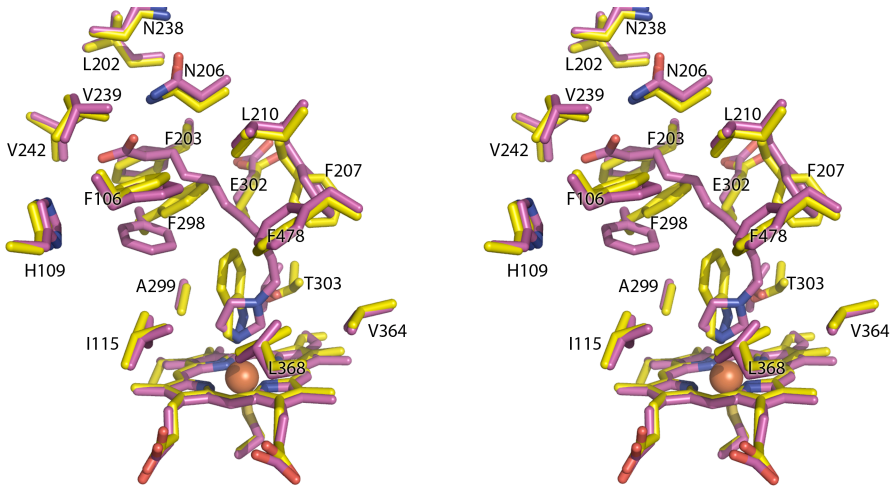


Figure 5.2: Changes in CYP2E1 active site residues in alternate ligand-bound structures shown in wall-eyed stereo view. Amino acid side chains within 5 Å of ω-imidazolyl-decanoic acid are shown. The ω-imidazolyl-decanoic acid-bound structure is shown in magenta while the indazole-bound structure is shown in yellow.

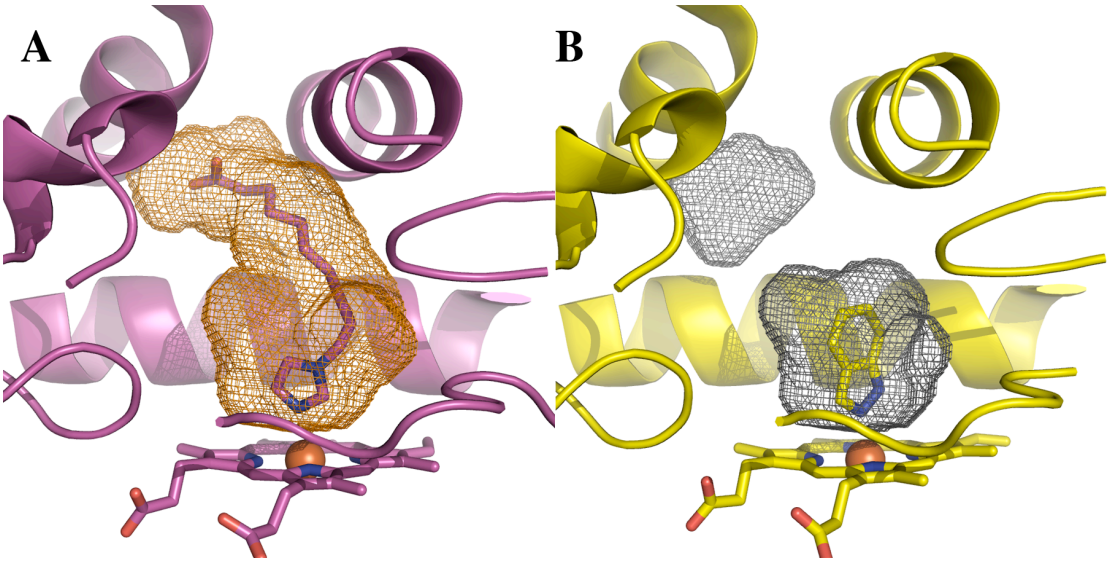


Figure 5.3: Comparison of CYP2E1 active site voids when bound with ω-imidazolyl-decanoic acid (**A**, active site volume of 440 $Å^3$) or indazole (**B**, active site volume of 190 $Å^3$ and 77 $Å^3$ for adjacent void) shown as orange and grey mesh respectively. Volumes calculated using the program VOIDOO (21).

Discussion

Ligand Selection

In order to successfully understand fatty acid binding, orientation, and positioning it was important to select ligands that would simultaneously satisfy several favorable properties. These properties include: compounds with high affinity for CYP2E1, inducement of spectral shift so the complex is easily identified, and moderate water solubility for cocrystallization. The ω-imidazolyl series of fatty acid analogs were closest to the fatty acid substrates metabolized by CYP2E1, satisfied the desired characteristics, and was easily generated through established synthetic protocols (5,8).

Fatty Acid Analog Binding Site

In previous work (Chapter 4) it was postulated that the residues of the conserved ^{216}QxxNN^{220} may act as hydrogen bond donors to orient fatty acid substrates with the ω- or (ω-1)-carbon in the active site for metabolism. The crystal structure described herein of CYP2E1 bound with the fatty acid analog ω-imidazolyl-decanoic acid describes a different binding mode for fatty acid interaction with CYP2E1. The CYP2E1 electron density map (omit σ_A-weighted $2|F_O| - |F_C|$ map contoured at 1.0 σ, Figure 5.1) shows continuous electron density from the imidazole ring of the decanoic acid analog to its carboxylate terminus. The imidazole nitrogen coordinates directly to the heme iron while the carboxylate hydrogen bonds to N206 in the F-helix.

The fatty acid analog cocrystallized with CYP2E1 is not the highest affinity compound examined in this series. It did, however, possess the necessary balance between water solubility and affinity to facilitate cocrystallization with CYP2E1. The binding data shows that the highest affinity ligand of the series is actually the ω-imidazolyl-dodecanoic acid. If ω-imidazolyl-dodecanoic acid binds in the same mode as ω-imidazolyl-decanoic acid, then the longer compound may have higher affinity because it may allow stronger interactions between that analog acid terminus and H109. H109 is opposite N206, the hydrogen bond donor in the CYP2E1/ω-imidazolyl-decanoic acid complex and is a distance of 4.3 Å from the current ligand's carboxylate. It should also be noted that just beyond H109 is K243. K243 has a distance of 10 Å from the current ligand's carboxylate. This series of three polar/positively charged amino acids (Figure 5.4 B) may allow for CYP2E1 to bind the broad range of fatty acid chain lengths that it is known to metabolize (24).

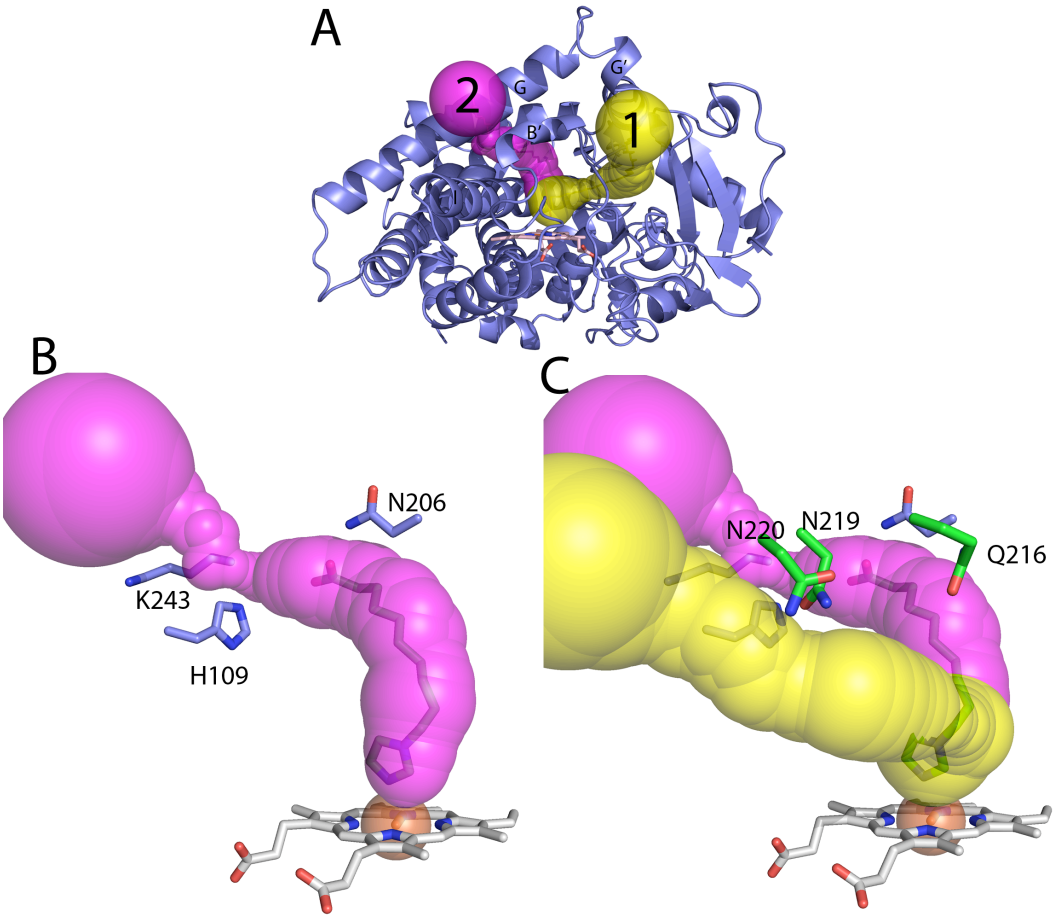


Figure 5.4: The top two scoring access channels of CYP2E1 (A) as calculated by CAVER (22). The yellow channel (path 1) was the highest scoring with magenta being second (path 2). The polar residues lining the magenta channel are N206, H109, and K243 (B). The polar residues lining the yellow channel are Q216, N219, and N220 (C).

Access Channel Calculations

The same calculations that were utilized in the analysis of the most likely route for ligand entry and egress in the CYP2E1/indazole structure (described in Chapter 4) were also used to probe the CYP2E1/ω-imidazolyl-decanoic acid structure. For the CYP2E1/indazole structure, CAVER (22) identified exit routes passing from the active site on either side of F478, through a large channel and ultimately exiting the protein between the B-B′ loop, the β_1 sheet system, and beneath the F′ and G′ helices (Figure 5.4 A, path 1). When the calculations were performed on the CYP2E1/ω-imidazolyl-decanoic acid structure the top result still exited the protein in the same position as previously described (Figure 5.4 A, path 1). However, the second highest scoring result exited the protein between the B′ and G-helices above the I-helix (Figure 5.4 A, path 2). This follows a similar tract to that of the VOIDOO calculated active site in the fatty acid analog structure (Figure 5.3 A). The environment outside of these portals may suggest the different roles of the two paths. Path 1 exits the protein close to the putative membrane bound portion of the protein while path 2 exits on a face of the protein that is expected to be more exposed to the cytosol (Figure 5.5). This may provide a path for hydrophobic substrates to enter the protein from the membrane via path 1, undergo metabolism in the active site, and exit through path 2 as the more hydrophilic metabolite.

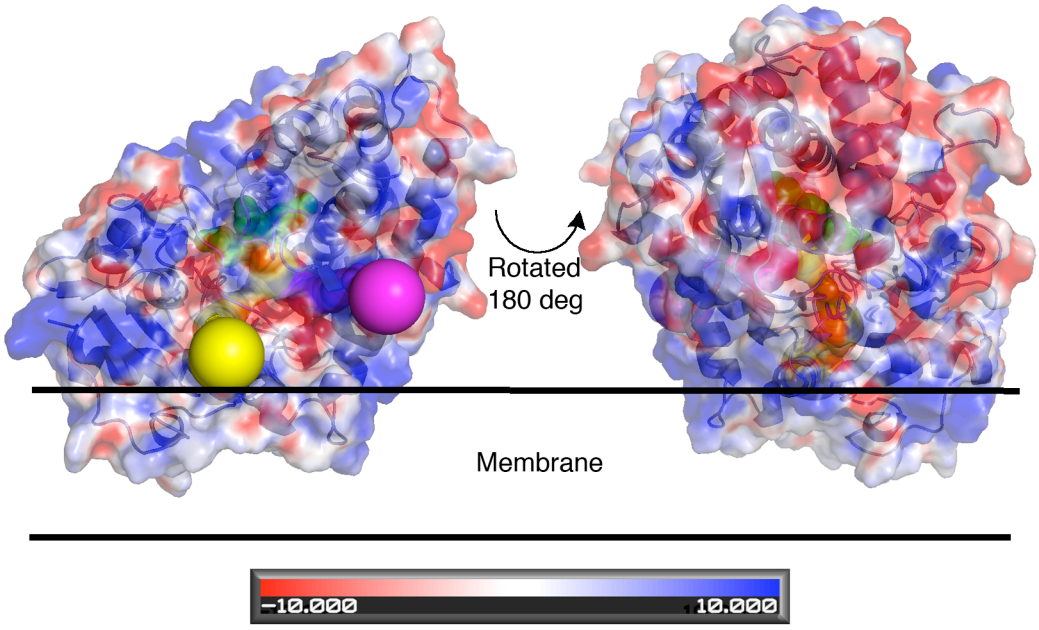


Figure 5.5: CAVER calculated cavities along with electrostatic surface. The proposed membrane binding region of the N-terminus, and F′ and G′ helices are at the bottom of this protein structures in this orientation. Two orientations are shown to illustrate the positively charged band of residues that would interact with polar head groups of the lipid bilayer. The colors of the CAVER paths correspond with Figure 5.4. For orientation the heme is shown as green spheres.

Fatty acid binding residues and their conservation among species

As discussed in Chapter 1, CYP2E1 is so conserved among mammals that the name is retained between species. Further, the CYP2E1 metabolism of fatty acids has been shown in several species using microsomal systems (9,25). This suggests that the amino acid residues responsible for the orientation of fatty acids may be conserved. The residue responsible for binding the carboxylate end of ω-imidazolyl-decanoic acid in this structure is N206. Further down exit path 2 (Figure 5.4 B) lie two additional positively charged amino acid side chains that may interact with the carboxylate of the longer fatty acids CYP2E1 is known to metabolize. These other potential H-bond donors are H109 and K243. Figure 5.6 shows sequence alignments of the regions of CYP2E1 containing N206, H109, and K243. All three of these residues are primarily conserved across species with these exceptions: in rat and mouse H109 is another H-bond donator Tyr, and in rabbit it is Phe; in mouse K243 is Arg.

Figure 5.6: CYP2E1 species sequence alignment. The primary sequence of human CYP2E1 from residues 61 to 300 were aligned with CYP2E1 from other mammalian species using CLUSTAL. The conserved residues involved / proposed to be involved in fatty acid binding are highlighted: H109 in blue with variations in gray; N206 in green; and K243 shown in yellow with variations in red. Species abbreviations: human-*Homo sapiens*, pig-*Sus scrofa*, whale-*Balaenoptera acutorostrata*, bovine-*Bos taurus*, rhesus-*Macaca mulatta*, marmos-*Callithrix jacchus*, rat-*Rattus norvegicus*, mouse-*Mus musculus*, rabbit-*Oryctolagus cuniculus*, dog-*Canis familiaris*, and cat-*Felis silvestris catus*.

```
P05181        AQRFGPVFTLYVGSQRMVVMHGYKAVKEALLDYKDEFSGRGDLPAFHAHRDRGIIFNNGP 120 human
Q8SQ64        AERYGPVFTVYLGSRRIVVLHGYKAVKEVLLHYKNEFSGRGEIPTFQVHKDKGVIFNNGP 120 pig
A6BM02        AERFGPVFTLYLGSRRFVVLHGYKAVKEVLLDYRNEFSGRGETPAFQVHQDKGIIFNNGP 114 whale
O18963        AERYGPVFTLYLGSQRAVVVHGYKPVKEVLLDYKNEFSGRGENPGFQMHKNNGIIFNNGS 120 bovine
Q6GUQ4        AQRFGPVFTLYVGSRRVVVVHGYKAVREVLLDHKDEFSGRGDIPAFHAHRDRGIIFNNGP 120 rhesus
Q6LEM3        AERFGPVFTLYLGARRVVVLYGYKAVREALLDYKSEFSGRGEIPAFREHKDRGIIFNNGP 120 marmos
P05182        AKRFGPVFTLHLGSRRIVVLHGYKAVKEVLLNHKNEFSGRGDIPVFQEYKNKGIIFNNGP 120 rat
Q05421        AKRFGPVFTLHLGQRRIVVLHGYKAVKEVLLNHKNEFSGRGDIPVFQEYKNKGIIFNNGP 120 mouse
P08682        AERFGPVFTVYLGSRRVVVLHGYKAVREMLLNHKNEFSGRGEIPAFREFKDKGIIFNNGP 120 rabbit
Q9MZY0        AEQYGPVFTLYLGSQRTVVLHGYKAVKEVLLDHKNDLSGRGEVFAFQSHKDRGITFNNGP 120 dog
Q5KR53        AERYGPVFTLYLGSQRTVVLHGYKAVKEVLLDYKNEFSGRGEIYAFEAHKDKGITFNNGP 120 cat

P05181        TWKDIRRFSLTTLRNYGMGKQGNESRIQREAHFLLEALRKTQGQPFDPTFLIGCAPCNVI 180
Q8SQ64        TWRDTRRFSLTTLRDFGMGKQGNEQRIQREAHFLLEALRKTHGQPFDPTFLIGCAPCNVI 180
A6BM02        TWQDTRRFSLTTLRDFGMGKQGNEQRIQSEAQLLLGALRKTHGQPFDPTFVIGFAPYNVI 174
O18963        TWRDTRRFSLTTLRDLGMGKQGNEQRIQREAHFLLEVLRKTQGQPFDPTFVVGFAPYNVI 180
Q6GUQ4        TWKDIRRFSLTTLRNYGMGKQGNESRIQREAHFLLEALRKTQGQPFDPTFLIGCAPCNVI 180
Q6LEM3        TWKDIRRFSLTALRNYGMGKQGNENRIQREAHFLVEALRKTQGQPFEPTFLIGCAPCNVI 180
P05182        TWKDVRRFSLSILRDWGMGKQGNEARIQREAQFLVEELKKTKGQPFDPTFLIGCAPCNVI 180
Q05421        TWKDVRRFSLSILRDWGMGKQGNEARIQREAHFLVEELKKTKGQPFDPTFLIGCAPCNVI 180
P08682        TWKDTRRFSLTTLRDYGMGKQGNEDRIQKEAHFLLEELRKTQGQPFDPTFVIGCTPFNVI 180
Q9MZY0        GWKDTRRLSLSTLRDYGMGKRGNEERIQREIPFLLEALRGTRGQPFDPTFLLGFAPFNVI 180
Q5KR53        SWKDTRRLSLSILRDYGMGKRANEEQIQREVPFLLEAFRGTQGQPFDPTFVLGYAPCNVI 180

P05181        ADILFRKHFDYNDEKFLRLMYLFNENFHLLSTPWLQLYNNFPSFLHYLPGSHRKVIKNVA 240
Q8SQ64        SDILFRQHFDYNDKTCLRLMSMFNENFYLLSTGWIQLYNNFSGYLRYLPGSHRKLMKNIS 240
A6BM02        SDILFHKRADYNDKTALRMLSLFNENFYLLSSPWIQLYNNFPGYIRYLPGSHRKLIKNVS 234
O18963        SDILFHKRFDYKDQTSLRLMSLFNENFYLLSSPWIQLYNNFPDYLQYLPGSHRKLLKNVS 240
Q6GUQ4        ADILFRKHFDYNDEKFLRLMYLFNENFQLLSTPWLQLYNNFPSLLHYLPGSHRKVMKNVA 240
Q6LEM3        ADILFRKRFDYDDEKFLRLMHLFNENFYLLSTPWLQLYNNFSTYLHYLPGSHRKVIRNVA 240
P05182        ADILFNKRFDYNDKKCLRLMSLFNENFYLLSTPWIQLYNNFADYLRYLPGSHRKIMKNVS 240
Q05421        ADILFNKRFDYDDKKCLELMSLFNENFYLLSTPWIQAYNYFSDYLQYLPGSHRKVMKNVS 240
P08682        AKILFNDRFDYKDKQALRLMSLFNENFYLLSTPWLQVYNNFSNYLQYMPGSHRKVIKNVS 240
Q9MZY0        ADILFHKHFDYSDQTGLRIQKLFNENFHLLSTGWLQLYNIFPSYLHYLPGSHRKVLRNVA 240
Q5KR53        SDILFRKRFDYRDQTGLRIQKLFNENFHLLSTNWLQLYNIFPSYMRYLPGNHRKVIKNVY 240

P05181        EVKEYVSERVKEHHQSLDPNCPRDLTDCLLVEMEKEKHSAERLYTMDGITVTVADLFFAG 300
Q8SQ64        EIKDYALERVKDHRDSLEPSCPRDFTDTLLMEMEKEKYSAEPIYTLDNIAVTVADMFFAG 300
A6BM02        EIKEYALEGVKDHQKSLEPSCPRDFTDTMLMEMEKEKHSTDPVYTLDNIAVTVADLLFAG 294
O18963        EVKSYALERVKDHQKSLEPSCPRGFLDTMLIEMAKERHSVDPMYTLENIAVTVADLLFAG 300
Q6GUQ4        EIKEYVSERVKEHLQSLDPNCPRDLTDCLLVEMEKEKHSAERLYTMDGITVTVADLFFAG 300
Q6LEM3        EIKEYVSERVKEHYQSLDPNCPRDLTDCLLVEMEKEKPSAEPLYTMDGITVTVADLFFAG 300
P05182        EIKQYTLEKAKEHLQSLDINCARDVTDCLLIEMEKEKHSQEPMYTMENVSVTLADLFFAG 300
Q05421        EIRQYTLGKAKEHLKSLDINCPRDVTDCLLIEMEKEKHSQEPMYTMENISVTLADLFFAG 300
P08682        EIKEYTLARVKEHHKSLDPSCPRDFIDSLLIEMEKDKHSTEPLYTLENIAVTVADMFFAG 300
Q9MZY0        ELKDYSLERVKEHQESLDPTCSRDFTDCLLQELQKERYGTEPWYTLDNIAVTVADLFFAG 300
Q5KR53        EIKSYTAARVKEHEESLDPNCPRDFTDSLLLEMRKERYTKEPWFTFDNIAATIADLFFAG 300
```

As previously mentioned, in addition to the metabolism of medium length fatty acids, CYP2E1 is known for the oxidation of eicosanoids (26,27). Two important arachidonic acid metabolites generated by CYP2E1 (along with other CYPs) are the 20-hydroxyeicosatraenoic acid (20-HETE) and 19-hydroxyeicosatraenoic acid (19-HETE). These metabolites are potent signaling molecules in that 20-HETE has been shown to be a vasoconstrictor (28) where 19-HETE has been shown to be a vasodilator (29). In previous work (Chapter 4), a binding mode for arachidonic acid was proposed along the large channel (Caver path 1). If arachidonic acid binds along path 2 like the fatty acid analog shown in this new complex, interactions of the arachidonic acid carboxylate with K243, could also position this substrate for ω- or (ω-1)-hydroxylation (Figure 5.7).

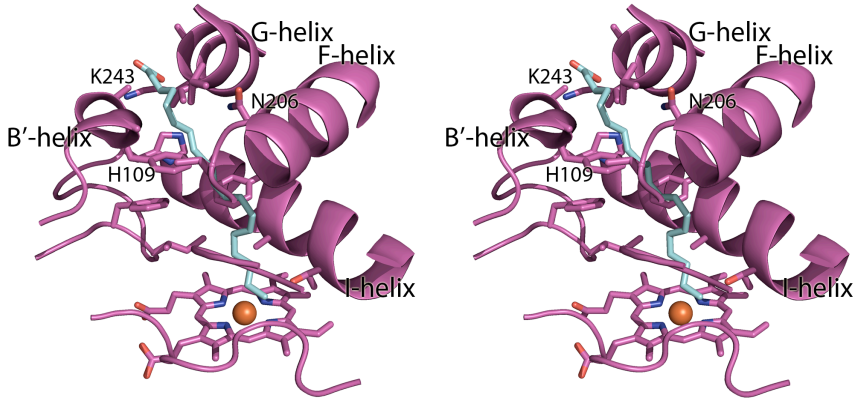


Figure 5.7: Arachidonic acid manually docked into the CYP2E1 structure, shown in wall-eyed stereo view. Arachidonic acid (shown in light blue) was manually docked into the substrate access channel path 2 positioning the arachidonic acid carboxylate for hydrogen bonding to K243 and opposing terminus for ω- or (ω-1)-hydroxylation.

Comparison of CYP2E1 and P450BM3 bound with fatty acid analogs

As previously discussed, an ω-imidazolyl fatty acid analog has been cocrystallized with P450BM3. To understand the similarities and differences between the CYP2E1 and P450BM3 complexes with fatty acid analogs the structures were superimposed using secondary structure matching (30). The alignment showed significant differences in the overall structure with a R.M.S.D. of the Cα atoms of 3.14 Å (Figure 5.8A). This is not surprising due to the low sequence identity (~20%) between the two proteins. Upon more detailed inspection the two proteins bind fatty acid analogs in very different orientations (Figure 5.8B). The carboxylates of these ligands oriented in opposite directions with respect to each other in the enzyme active sites. Both enzymes catalyze the sub-terminal hydroxylation of fatty acids (CYP2E1 mainly ω-1 and P450BM3 mainly ω-2 but also ω-1 and ω-3 (31)) at similar positions along with the formation of epoxides on unsaturated fatty acids (32,33). While the relationship of these proteins is an example of convergent functions, the structures are divergent.

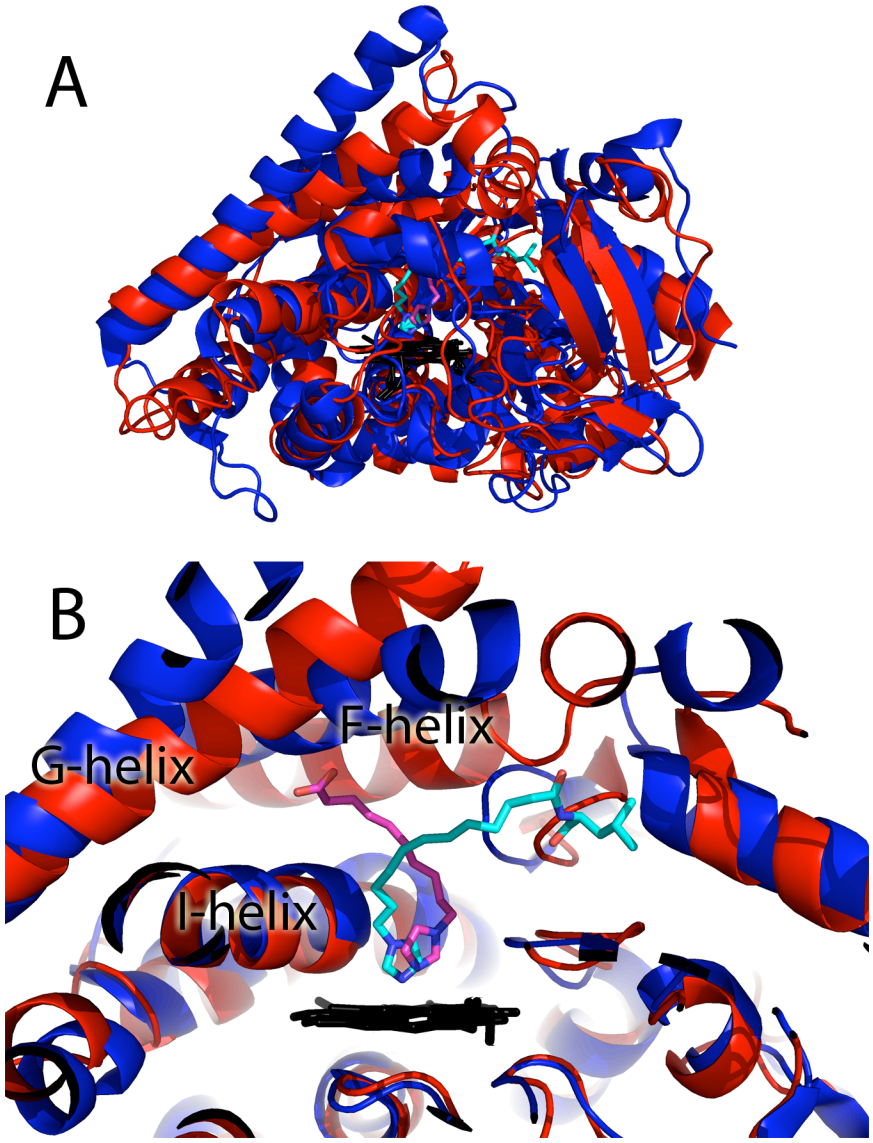


Figure 5.8: Structural overlay of CYP2E1 and P450BM3 bound with fatty acid analogs. The structures of CYP2E1 (red, heme in black) bound with ω-imidazolyl-decanoic acid (magenta) and P450BM3 (blue, heme in black) bound with *N*-(10-imidazolyldodecanoyl)-*L*-leucine (teal) were superimposed using secondary-structure matching (30) in COOT (17) with a Cα root mean square deviation of 3.14 Å. There are considerable differences in the overall structure (A) contributing the different ligand binding modes shown in the active site cross section (B).

Conclusions

In summary, the structure of CYP2E1 bound to the fatty acid analog, ω-imidazolyl-decanoic acid, has shown that subtle active site changes can allow the active site volume to more than double to accommodate fatty acids. This active site flexibility opens a different portion of the protein for fatty acid binding than previously proposed. In addition, it opens another location for potential access/egress to/from the active site. Future work is needed to fully understand how CYP2E1 acclimates to even larger fatty acids such as arachidonic acid, whose metabolism has a significant physiological impact.

References

1. Kroetz, D. L., Yook, P., Costet, P., Bianchi, P., and Pineau, T. (1998) *J Biol Chem* **273**(47), 31581-31589
2. Haines, D. C., Tomchick, D. R., Machius, M., and Peterson, J. A. (2001) *Biochemistry* **40**(45), 13456-13465
3. Joyce, M. G., Girvan, H. M., Munro, A. W., and Leys, D. (2004) *J Biol Chem* **279**(22), 23287-23293
4. Haines, D. C., Chen, B., Tomchick, D. R., Bondlela, M., Hegde, A., Machius, M., and Peterson, J. A. (2008) *Biochemistry* **47**(12), 3662-3670
5. Alterman, M. A., Chaurasia, C. S., Lu, P., Hardwick, J. P., and Hanzlik, R. P. (1995) *Arch Biochem Biophys* **320**(2), 289-296
6. Bambal, R. B., and Hanzlik, R. P. (1996) *Arch Biochem Biophys* **334**(1), 59-66
7. Chaurasia, C. S., Alterman, M. A., Lu, P., and Hanzlik, R. P. (1995) *Arch Biochem Biophys* **317**(1), 161-169
8. Lu, P., Alterman, M. A., Chaurasia, C. S., Bambal, R. B., and Hanzlik, R. P. (1997) *Arch Biochem Biophys* **337**(1), 1-7
9. Fukuda, T., Imai, Y., Komori, M., Nakamura, M., Kusunose, E., Satouchi, K., and Kusunose, M. (1994) *J Biochem* **115**(2), 338-344
10. Moran, J. H., Mitchell, L. A., Bradbury, J. A., Qu, W., Zeldin, D. C., Schnellmann, R. G., and Grant, D. F. (2000) *Toxicology and applied pharmacology* **168**(3), 268-279

11. Porubsky, P. R., Meneely, K. M., and Scott, E. E. (2008) *J Biol Chem* **283**(48), 33698-33707
12. Bookser, B. C., and Zhu, S. (2001) *J. Comb. Chem* **3**(2), 205-215
13. Luft, J. R., and DeTitta, G. T. (1999) *Acta Crystallographica Section D* **55**(5), 988-993
14. Rossmann, M. G., and van Beek, C. G. (1999) *Acta Crystallogr D Biol Crystallogr* **55**(Pt 10), 1631-1640
15. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In. *Macromolecular Crystallography, Pt A*, Academic Press Inc, San Diego
16. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40**(4), 658-674
17. Emsley, P., and Cowtan, K. (2004) *Acta crystallographica* **60**(Pt 12 Pt 1), 2126-2132
18. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta crystallographica* **53**(Pt 3), 240-255
19. Vriend, G. (1990) *J. Mol. Graphics* **8**(1), 52-56, 29
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**(2), 283-291
21. Kleywegt, G. J., and Jones, T. A. (1994) *Acta crystallographica* **50**(Pt 2), 178-185
22. Petrek, M., Otyepka, M., Banas, P., Kosinova, P., Koca, J., and Damborsky, J. (2006) *BMC bioinformatics* **7**, 316
23. Krissinel, E., and Henrick, K. (2004) *Acta Crystallographica Section D* **60**(12 Part 1), 2256-2268
24. Adas, F., Salaun, J. P., Berthou, F., Picart, D., Simon, B., and Amet, Y. (1999) *Journal of lipid research* **40**(11), 1990-1997
25. Amet, Y., Berthou, F., Goasduff, T., Salaun, J. P., Le Breton, L., and Menez, J. F. (1994) *Biochem Biophys Res Commun* **203**(2), 1168-1174
26. Laethem, R. M., Balazy, M., Falck, J. R., Laethem, C. L., and Koop, D. R. (1993) *J. Biol. Chem.* **268**(17), 12912-12918
27. Moran, J. H., Mitchell, L. A., Bradbury, J. A., Qu, W., Zeldin, D. C., Schnellmann, R. G., and Grant, D. F. (2000) *Toxicology and applied pharmacology* **168**(3), 268-279
28. Ma, Y. H., Gebremedhin, D., Schwartzman, M. L., Falck, J. R., Clark, J. E., Masters, B. S., Harder, D. R., and Roman, R. J. (1993) *Circ Res* **72**(1), 126-136
29. Escalante, B., Falck, J. R., Yadagiri, P., Sun, L. M., and Laniado-Schwartzman, M. (1988) *Biochem Biophys Res Commun* **152**(3), 1269-1274
30. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 12 Pt 1), 2256-2268
31. Lewis, D. F. V. (2001) *Guide to Cytochromes P450 Structure and Function*, Taylor and Francis, London and New York

32. Capdevila, J. H., Wei, S., Helvig, C., Falck, J. R., Belosludtsev, Y., Truan, G., Graham-Lorence, S. E., and Peterson, J. A. (1996) *J Biol Chem* **271**(37), 22663-22671
33. Celik, A., Sperandio, D., Speight, R. E., and Turner, N. J. (2005) *Org Biomol Chem* **3**(15), 2688-2690

Page left intentionally blank

Chapter 6.

Conclusions

The work described in this thesis focused on the two CYP isoforms: CYP2A13 and CYP2E1. These enzymes are important because of their involvement in xenobiotic metabolism and the associated subsequent toxicity. Understanding the individual active site topology of these xenobiotic-metabolizing enzymes is important to understand both their pharmacological and toxicological roles.

In the first crystal structure of CYP2A13, density of unknown origin was found in the enzyme active site. Several factors suggested that the ligand might be indole or an indole metabolite so an existing method for detecting indole in bacterial culture was modified to a more sensitive HPLC-UV/Vis-MS/MS method that could differentiate between indole and several oxidized forms of indole. The method was used to analyze protein used for crystallization and definitively identified unsubstituted indole in the solution, leading to its subsequent modeling into the CYP2A13 active site electron density.

The first crystal structure of CYP2E1 was solved with two different ligands bound. Initially the structure of CYP2E1 was solved with the small molecular weight inhibitor indazole. This structure showed a protein with the typical P450 fold. The active site was compact (190 $Å^3$), hydrophobic, and smaller than any other CYP structurally characterized. This active site morphology was consistent with the CYP2E1 role in metabolism of small molecular weight substrates. Directly adjacent to the active site lies a channel that extends to the enzyme surface that has been

proposed to allow access of substrates to the CYP catalytic core. An additional void with a volume of 77 $Å^3$ is distal to the active site cavity with the residues of F106, F116, and F298 separating the two voids.

CYP2E1 is also known to metabolize fatty acids and it was still unclear what active site conformational changes were necessary to bind the larger substrates. It was postulated from the indazole-bound CYP2E1 structure that fatty acids could extend from a group of polar amino acid in the substrate access channel into the active site to position the ω-teminus of the fatty acid for ω-1 metabolism. To investigate the validity of this proposal, CYP2E1 was cocrystallized and the structure solved with a fatty acid analog, ω-imidazolyl-decanoic acid. The movement of F298 allowed merging of the active site and the small adjacent empty cavity increasing the active site volume to 440 $Å^3$. The carboxylate of this analog hydrogen bonds directly to N209 of the F-helix, which is far from the previously proposed polar amino acids of the substrate access channel. This active site flexibility may also allow for the binding of the larger fatty acids (like arachidonic acid) that CYP2E1 is also known to metabolize.

This research provides important new information about ligand binding in two human CYP enzymes that will be useful in predicting CYP metabolism, designing drugs that are resistant to CYP metabolism, and designing pro-drugs that will be metabolized by these enzymes. In the case of CYP2E1, the structure can be used as a basis for the design of selective inhibitors to prevent the toxicity of acetaminophen overdose. Future work should focus on confirming the binding modes of larger fatty

acid molecules, like arachidonic acid, whose metabolites play important physiological roles.